

**(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)**

**(19) World Intellectual Property Organization**  
International Bureau



**(43) International Publication Date**  
**13 June 2002 (13.06.2002)**

**PCT**

**(10) International Publication Number**  
**WO 02/46444 A2**

- (51) International Patent Classification<sup>7</sup>:** **C12N 15/90**
- (21) International Application Number:** PCT/US01/46311
- (22) International Filing Date:** 5 December 2001 (05.12.2001)
- (25) Filing Language:** English
- (26) Publication Language:** English
- (30) Priority Data:**  
60/251,482 5 December 2000 (05.12.2000) US
- (71) Applicant:** **WISCONSIN ALUMNI RESEARCH FOUNDATION** [US/US]; 614 North Walnut Street, P.O. Box 7365, Madison, WI 53707-7365 (US).
- (72) Inventors:** **GORYSHIN, Igor, Yu**; Apt. D, 23 University Houses, Madison, WI 53705 (US). **NAUMANN, Todd, A.**; 2233 Concord Drive, State College, PA 16801 (US). **REZNIKOFF, William, S.**; 733 Lakewood Boulevard, Madison, WI 53705 (US).
- (74) Agent:** **BERSON, Bennett, J.**; Quarles & Brady LLP, P.O. Box 2113, Madison, WI 53701-2113 (US).
- (81) Designated States (national):** AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.
- (84) Designated States (regional):** ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

**Published:**

- without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.



**WO 02/46444 A2**

- (54) Title:** DOUBLE TRANSPOSITION METHODS FOR MANIPULATING NUCLEIC ACIDS

**(57) Abstract:** Methods are provided for manipulating nucleic acid to produce gene fusions, to delete or clone a portion of a chromosome, or to insert a sequence into a chromosome. The methods employ sequential transposition processes using two or more pairs of inverted repeat transposase-interacting sequences on a transposable polynucleotide wherein each pair of transposase-interacting sequences interacts with a distinct transposase enzyme.

## DOUBLE TRANSPOSITION METHODS FOR MANIPULATING NUCLEIC ACIDS

## CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of US Provisional Patent application number 60/251,482, filed December 5, 2000, which application is incorporated by reference herein as if set forth in its entirety.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH  
OR DEVELOPMENT

[0002] This invention was made with United States government support awarded by the National Institutes of Health, Grant No. GM50692. The United States has certain rights in this invention.

## BACKGROUND OF THE INVENTION

[0003] Manipulation of nucleic acids and proteins is an important aspect of modern molecular biology. In particular, the science of combinatorial genetics has advanced in recent years as it has become apparent that proteins having altered structure and function can be engineered by swapping large or small portions of the amino acid sequence with other related or unrelated amino acid sequences. Using this approach, it is also possible to engineer novel proteins that bring together in a single molecule the structures and functions of diverse molecules. Such manipulations are most readily undertaken at the nucleic acid level. The nucleic acids thus produced can be transcribed to produce fusion RNAs and translated either *in vitro* or *in vivo* using known methods and the recombinant proteins thus produced can be isolated. Other manipulations, such as inserting polynucleotides of interest into a chromosome, deleting sections of a chromosome, or cloning sections of a chromosome are also of interest. Moreover, various approaches are known for shuffling gene pieces and selecting or screening for products having one or more desired activities or properties. Examples of such technologies include US Patent Numbers 5,605,793, 5,830,721, and 6,132,970. A number of companies including Maxygen, Diversa, Applied Molecular Evolution, and Genencor International among others specialize in directed molecular evolution. See Pollack, A., "Selling Evolution in Ways Darwin Never Imagined," New York Times, page B1, October 28, 2000.

[0004] Recently, non-transposition based methods for generating large libraries of randomly fused genes have been reported. Ostermeir and co-workers first described a method

termed ITCHY that involves the incremental truncation of two genes by use of ExoIII nuclease followed by S1 nuclease treatment, polymerization, and ligation of fragments to form random fusions (Ostermeier *et al.*, 1999; Lutz *et al.*, 2000). A second method has also been reported that utilizes random cleavage of DNA followed by a series of digestion and ligation reactions to create random fusions. This technique, called SHIPREC, also has features that increase the amount of useful fusions if both proteins have similar length and domain organization (Seiber *et al.*, 2001). Existing methods for gene shuffling and chromosome manipulation are complex and can exhibit a bias for or against particular recombination sites and sequences. Thus, the art continues to develop more sophisticated manipulations.

[0005] Various aspects of an *in vitro* transposition system that employs sequences from, and sequences derived from, the Tn5 transposon are described in US Patent Numbers 5,925,545, 5,948,622, and 5,965,443, each of which is incorporated by reference herein as if set forth in its entirety. International publication Number WO 00/17343, also incorporated herein by reference as if set forth herein in its entirety, discloses a system for introducing into cells synaptic complexes that comprise a Tn5 transposase and a polynucleotide having flanking sequences that operably interact with the transposase to form a synaptic complex.

[0006] Even though these known systems for Tn5-based *in vitro* transposition are effective and very useful, they do not provide sufficient manipulative control to meet the technological goals noted above.

[0007] Efforts are also underway to define so-called minimal bacterial genomes for growth under defined conditions and, similarly, to identify genes essential for growth under defined conditions. Determining the content required for a minimal bacterial genome is of intense interest. One approach is to assemble the theoretical minimal genome *in silico* by comparing a variety of different microbial genomes. Alternatively, the smallest genome amongst existing genomes (mycoplasma) can be analyzed by mutagenesis. *E.coli* K12 is a preferred bacterium, because of its simplicity in handling, and its short generation time. It is desirable to try to generate a minimal or significantly reduced *E.coli* K12 genome, which may shorten the already short doubling time in rich media.

[0008] Recently developed transposon-based approaches involve inserting a transposon into a gene to (1) knockout or disrupt a gene function or (2) introduce a lethal mutation that cannot be observed in an essential gene. These methods essentially catalogue transposition into non-essential genes. It is assumed that any gene that contains no transposon insert is essential.

[0009] An important alternative approach involves affirmatively identifying essential genes in libraries of cells, where the cells contain transposons having selectively regulated outwardly-facing promoters inserted upstream from an essential gene. While the cells of interest are not viable on media that cannot activate a transposon promoter, expression is restored by selectively activating either or both of the transposon promoters. Unfortunately, only a few of the transposon inserts in a library will insert into the promoter region of an essential gene.

#### BRIEF SUMMARY OF THE INVENTION

[0010] In a first aspect, the invention is summarized in that a transposable polynucleotide is suitable for use in methods for manipulating nucleic acids to create libraries of cells that contain transposed nucleic acid when the polynucleotide comprises two or more inverted repeat sequence pairs (“transposase-interacting sequences”) arranged as disclosed herein, where each pair has a distinct and separable ability to interact with a distinct transposase enzyme. The pairs can be provided in a nested fashion such that both members of one pair are flanked by both members of the second pair. In a related aspect, it is not essential to provide two complete pairs of transposase-interacting sequences in a single transposon, nor to introduce all of the transposase-interacting sequences by transposition. Rather, as will become more apparent below, a first member of one pair can separately be provided directly on the substrate DNA for use in a second of two transposition events with a second member of the pair, the second member being introduced during a first of two transposition events. By providing and arranging selectable markers, origins of replication, and/or other nucleotide sequences of interest on the transposable polynucleotide, one can direct sequential transposition processes to achieve the desired manipulated nucleic acids.

[0011] In a related aspect, the invention is a transposable polynucleotide having transposase-interacting sequences and additional sequences arranged as disclosed elsewhere herein. The nature of the arrangement dictates the uses to which the transposable polynucleotides can be put.

[0012] In yet another aspect, the invention is summarized in that the transposable polynucleotides are introduced into host cells as disclosed to produce libraries of cells that achieve the objects of the invention. Using selection and screening methods, cells that have undergone desired transpositions can be obtained. In certain embodiments, the host cells transcribe and translate double transposition products produced entirely *in vitro*. In other embodiments, a synaptic complex formed *in vitro* between a first transposase and a transposable

polynucleotide of the invention is introduced into a host cell whereupon the transposable polynucleotide is transposed into the host cell chromosome. This first transposition is followed, upon induction of a second transposase, by a second round of transposition *in vivo* to yield chromosomal insertions or deletions or cloned products excised from the host cell chromosome.

[0013] In another aspect, the invention is summarized in that a host cell contains an extrachromosomal circular polynucleotide that comprises first and second polynucleotide fusion portions, each polynucleotide fusion portion comprising a pair of segments from a pair of sequences that can encode a fusion RNA that can be translated to produce a fusion polypeptide, the members of the pair flanking a pair of transposase-interacting sequences each having a distinct and separable ability to interact with one of a pair of transposase enzymes, the pair of transposase-interacting sequences defining a junction between the two segments. In certain embodiments, the junction has a nucleic acid sequence that permits transcription and translation of both fusion segments across the junction.

[0014] In yet another aspect, the present invention is summarized in that a host cell contains a chromosome deleted in length relative to a wild type chromosome and comprising at the deletion site in place of the excised material, a pair of transposase-interacting sequences each having a distinct and separable ability to interact with one of a pair of transposase enzymes. In certain embodiments, the host cell further contains a self-replicating nucleic acid molecule that itself comprises the deleted portion of the host cell chromosome. In other embodiments, the host cell chromosome further comprises at the deletion site a segment of pre-selected nucleic acid between the pair of transposase-interacting sequences.

[0015] Still another aspect of the invention is summarized in that the transposase-interacting sequences are inverted repeats provided in inverted pairs on the polynucleotide for transposition. The two kinds of transposase-interacting sequences can be separate from one another, can be abutted end to end, or can overlap when it is particularly desirable to minimize the size of such sequence elements. As is detailed below, the transposase-interacting sequences can be engineered to ensure an open reading frame in either or both polarity, as desired.

[0016] It is an object of the invention to provide a system for manipulating nucleic acids *in vitro* and *in vivo*.

[0017] It is another object of the invention to generate a library of random fusions between two polynucleotide or polypeptide sequences, preferably between sequences that, respectively, encode and define peptides or proteins of interest.

[0018] It is a feature of the invention that the system employs two pairs of transposase-interacting sequences each interacting with a distinct transposase.

[0019] It is an advantage of the invention that transpositions occur without regard to the sequences of the nucleic acids into which the transposable elements transpose.

[0020] It is another advantage of the invention that large libraries (e.g.,  $>10^7$  individuals) having a high level of variability can be produced.

[0021] Other objects, features and advantages will become apparent upon consideration of the following detailed description taken in conjunction with the accompanying drawings.

#### BRIEF DESCRIPTION OF THE SEVERAL VIEWS OF THE DRAWINGS

[0022] Fig. 1 depicts a pair of preferred arrangements of full-length and interleaved (compressed) transposase-binding sequences that face in opposite directions.

[0023] Fig. 2 depicts a first approach to a method for producing a gene fusion library for obtaining directed gene and protein evolution products. In the first approach, a first transposase binds to and interacts with first transposase-interacting sequences in a donor polynucleotide (shown schematically as black triangles) whereupon the newly created transposon is transposed into a first target nucleic acid molecule that contains an appropriately-oriented transposase-interacting sequence to yield a set of first transposition products. Then, a second transposase binds to and interacts with the second transposase-interacting sequences, now both resident in the members of the set of first transposition products (shown schematically as white triangles), whereupon the donor portions of the first products are transposed into a second target nucleic acid molecule. The resulting nucleic acid molecule comprises a fusion sequence having a junction portion that includes a pair of distinct transposase-interacting sequences. It will be understood that only a few exemplary transposition products are shown, but that, in fact, the library of transposition products produced reflects the combinatorial variety resulting from the sequential transposition processes.

[0024] Fig. 3 depicts a second approach to a method for producing a gene fusion library for obtaining directed gene and protein evolution products. In the second approach, a first transposase binds to and interacts with first transposase-interacting sequences in a donor polynucleotide (shown schematically as black triangles) whereupon the newly created transposon is transposed into a first target nucleic acid molecule to yield a set of first transposition products. Then, a second transposase binds to and interacts with the second transposase-interacting sequences in the first transposition products (shown schematically as white triangles) whereupon

the donor portions of the first products are transposed into a second target nucleic acid molecule. The resulting nucleic acid molecule comprises a pair of fusion sequences, each fusion sequence having a junction portion that includes a pair of distinct transposase-interacting sequences. Again, it will be understood that only a few exemplary transposition products are shown, but that, in fact, the library of transposition products produced reflect the combinatorial variety resulting from the sequential transposition processes.

[0025] Fig. 4 depicts a first aspect of a construct and method for inserting the construct into a chromosome of bacterial host cell using the method for incorporating a synaptic complex between a transposase enzyme and donor DNA into a host cell chromosome as disclosed in incorporated International patent Application No. WO 00/17343. Other constructs described below can also be inserted into a bacterial host cell chromosome using this same first aspect.

[0026] Fig. 5 depicts intrachromosomal *in vivo* transposition of a construct introduced into the bacterial chromosome as generally shown and described in Fig. 4. This *in vivo* transposition yields a deleted chromosome carrying a small residual mismatched pair of transposase binding sequences. A second circular product carries a segment of deleted chromosomal material that cannot replicate because it lacks a bacterial origin of replication.

[0027] Fig. 6 depicts a screen for deletion products following the double transposition method of Figs. 4 and 5.

[0028] Fig. 7 depicts intrachromosomal *in vivo* transposition of a construct introduced into the bacterial chromosome as generally shown and described in Fig. 4. By including an origin of replication in this construct, the deleted chromosomal material (as in Fig. 5) can be maintained, amplified, and isolated for further use or study.

[0029] Fig. 8 depicts an intrachromosomal *in vivo* transposition of a construct introduced into the bacterial chromosome as generally shown and described in Fig. 4. By including in this construct a desired insert, between a mismatched pair of transposase binding sequences, the resulting bacterial chromosome includes the insert at the site of the chromosomal deletion that yields a desired insert added into the bacterial chromosome at the site of the first transposition event.

[0030] Fig. 9 depicts suitable transposons Tn5DEL7 and Tn5DEL8 for forming, respectively, deletions in a bacterial chromosome, and deletions in a bacterial chromosome while capturing the deleted portions in an extrachromosomal plasmid.

[0031] Fig. 10 depicts a set of bacterial chromosomal deletions induced in a method according to the invention.

## DETAILED DESCRIPTION OF THE INVENTION

[0032] Mutant Tn5 transposase enzymes can preferentially bind to and interact with distinct inverted repeat sequences. The end preference of a mutant transposase can be characterized either (1) by the *in vivo* transposition frequency observed when it is used in a system in which the target polynucleotide is flanked with particular termini, or (2) by the ratio of the *in vivo* transposition frequencies observed when it is used in a pair of systems in which the target polynucleotide is flanked with the termini, respectively. *In vitro* transposition frequency characterization is also possible.

[0033] The constructs and methods of the invention all employ two pairs of inverted repeat sequences (“transposase-interacting sequences”) that bind to and interact with a pair of transposase enzymes to facilitate directed double transposition as detailed below. The transposases are referred to generally herein as transposase 1(Tnp1) and transposase 2(Tnp2). There are no preordained requirements for the transposase-interacting sequences except that the combination of Tnp1 and the sequences that preferentially bind to and interact with it have a distinct and separable activity from the combination of Tnp2 and its preferential binding sequences.

[0034] By way of example, the Tn5 transposon includes inside end sequences (IE) and outside end sequences (OE). A transposase that preferentially binds to and interacts with OE, but not IE, is known and disclosed in incorporated US Patent Number 5,965,443. The hyperactive transposase there disclosed includes a mutation at position 54 and a mutation at position 372 that confer upon the Tn5 transposase a greater avidity for Tn5 outside end repeat sequences than wild type Tn5 transposase. A preferred embodiment of that transposase includes a lysine at position 54 and a proline at position 372. A most preferred embodiment of that transposase also includes an alanine residue at position 56. The preferred transposase of US 5,965,443 is referred to as Tnp EK/LP. All embodiments of the transposase disclosed in the incorporated patent are useful in the present invention. In the examples that follow, this transposase is represented as Tnp2.

[0035] Further incorporated US Patent No. 5,925,545 discloses that a TnpEK/LP-transposase-catalyzed *in vitro* transposition frequency at least as high as that of wild type OE is achieved if the termini include bases ATA at positions 10, 11, and 12, respectively, as well as the nucleotides in common between wild type OE and IE (e.g., at positions 1-3, 5-9, 13, 14, 16, and optionally 19). The nucleotides at positions 4, 15, 17, and 18 can correspond to the nucleotides found at those positions in either wild type OE or wild type IE. It is noted that the transposition

frequency can be further enhanced over that of wild type OE if the nucleotide at position 4 is a T. The so-called “mosaic end sequences” (“ME”) of US Patent No. 5,925,545 can be advantageously used in the present invention instead of wild-type OE sequences. Where “ME” sequences are used in the patent application, it is understood that ME and wild type OE sequences can be used, although use of ME yields a higher transposition frequency. Accordingly, Tnp2 can also be referred to as Tnp-OE or Tnp-ME. It is noted that, in the presence of Tnp EK/LP (Tnp2), a transposon having one ME and one OE transposes at higher frequency than one having a pair of OE termini. Transposons having ME-ME termini seem to transpose at a still slightly higher frequency.

[0036] On the other hand, a second class of hyperactive Tn5 transposase mutants exhibits a dramatic preference for Tn5 inside end sequences, and in some cases, methylated inside ends (IE<sup>ME</sup>), rather than for outside ends (OE), which are unmethylated because they lack a methylation site. The members of this class of Tn5 transposase mutants differ from wild-type Tn5 transposase protein at least at amino acid position 58 or position 372. In a preferred version, the second hyperactive Tn5 transposase mutant differs from wild-type in that it includes a cysteine at amino acid position 8, a valine at amino acid position 58, a lysine at amino acid position 344, and a glutamine at amino acid position 372. This transposase having four differences from wild type Tn5 transposase is referred to as Tnp sC7v2.0 and is used in the examples that follow, where it is represented as Tnp1 or Tnp-IE.

[0037] The aforementioned transposase enzymes having distinct and separable abilities to interact with distinct transposase-interacting sequences are preferred transposases for use in the methods of the invention. However, as long as one coordinates the relative positions of the transposase-interacting sequence pairs with the sequence (order) of providing the appropriate transposase, there is no requirement that either one of the transposases corresponds to transposase Tnp1 or Tnp2 as shown in the Figures. It is advantageous that the transposase enzymes have a hyperactive transposition efficiency relative to wild-type Tn5 transposase. Particularly when screening for transposition events, it is desirable for every cell to have undergone a transposition event. Accordingly, as a guideline, at least a 1000 fold activity increase relative to wild-type Tn5 transposase is considered to be preferred. The Tnp sC7v2.0 IE-specific enzyme described above is about 1000 fold more active than wild-type Tn5 transposase. The Tnp EK/LP is at least again about 10-100 fold more hyperactive than Tnp sC7v2.0. Mutant derivatives of these enzymes having still higher transposition activities while retaining their indicated specificities would be advantageously employed in the methods.

[0038] One can practice the method by providing the pairs of transposase-interacting sequences on a transposon in a nested manner such that the outer pair is employed to introduce the transposon into a target DNA. As an alternative, one member of a first pair of transposase-interacting sequences can be provided on the target while the second is provided in a proper orientation on the transposon, flanked by the second pair. When the transposon is subsequently introduced into the target, the first and second members of the first pair of transposase-interacting sequences are then positioned for subsequent transposition in the presence of a suitable transposase enzyme.

[0039] It is also possible, but less preferred, to practice the methods of the invention using a single transposase enzyme with pairs of the corresponding end sequence arranged in inverted back-to-back positions. In this approach, one can rely upon the screening and selection methods to yield only the transposition products of interest. However, the method is less preferred because it cedes some measure of control over the transposition processes themselves.

[0040] The methods of the invention are described below and are conceptually related in that each takes advantage first of one transposase/end sequence set to accomplish a first transposition event and then, second, of a second transposase/end sequence set to accomplish the second transposition event. By carefully engineering the placement of various components on the transposable element in each case, one can produce fusion RNAs, fusion proteins, chromosomal deletions, chromosomal insertions, or cloned chromosomal segments. The precise structures of the pairs of transposase-interacting sequences are described below. At this juncture, it is important only to note when considering the following disclosure, that, in this application, Tnp1 preferentially binds to and interacts with the end sequences shown schematically as black triangles while Tnp2 preferentially binds to and interacts with the end sequences shown as white triangles. The triangle points signify the inverted orientations of the members of each pair.

[0041] In the methods of the present invention, Tnp1 and Tnp2 are separately and sequentially provided to donor and target DNA molecules to promote directed evolution by two sequential transposition events. The result of the various methods disclosed depends upon the components provided on the transposed polynucleotides. Conceptually, it makes no difference whether Tnp1 or Tnp2 is used first, as long as the ends are provided in an orientation that directs the two transposition events to occur in the desired order.

[0042] There is some flexibility permitted in designing the sequences of the transposase-interacting sequences without appreciably reducing the OE- (or ME-) or IE-specific binding of the transposases. The incorporated US patents describe which bases of the end sequences can be

altered without unacceptably reducing the hyperactivity of the appropriate transposase. The ability to modify the end sequences can be advantageously exploited to produce a linker, such as the linker shown as compressed Linker A in Fig. 1, wherein the OE-related sequence (OE') and the IE-related sequence (IE') overlap with one another in a linker that is shorter than would be achieved by abutting the OE and IE sequences to one another. Linker A was designed by combining the IE and OE sequences. The overlap region does not match exactly the sequence of either IE or OE, hence the designations IE' and OE'. The 19 base pairs read from the top left are the IE sequence with position 18 changed from C to G (IE'). The 19 base pairs read from the bottom right match the OE sequence with positions 17 and 18 (linker positions 16 and 15) changed from AG to TC (OE'). As an alternative, compressed Linker B in Fig. 1 depicts a similar overlapped dual function linker engineered to respond to IE and ME, the mosaic ends mentioned above. Compressed Linker B differs from compressed Linker A only at position 29, as shown.

[0043] The full-length version of an IE/OE linker of Fig. 1 is accorded SEQ ID NO:1. The compressed version of an IE/OE linker of Fig. 1 (Linker A) is accorded SEQ ID NO:2. The full-length version of an IE/ME linker of Fig. 1 is accorded SEQ ID NO:3. The compressed version of an IE/ME linker of Fig. 1 (Linker B) is accorded SEQ ID NO:4.

[0044] The ability to translate gene B - gene A and gene A - gene B fusion proteins depends upon the presence in the linker region of an open reading frame in one or both directions. By carefully designing the linkers, one can allow both fusion products to be translated from start to finish or can ensure that only one fusion orientation is functional. The linkers can also be engineered to ensure that at least one reading frame is free of stop codons. This attribute ensures that combinatorial fusion proteins are translated through the transposition site. It may be preferable to design the linker so that only one reading frame is open so that all functional fusions have the same amino acids inserted at the transposition junction. It is also possible to recover functional fusion proteins in the method by altering either or both of the gene A and gene B starting target sequences such that no functional protein can be produced without two sequential transposition events, as described.

[0045] When producing gene fusion products according to the invention, the length of the linker is also important. If the length of the linker is a multiple of three and the linker contains only one open reading frame, the fusion point between the protein segments will have a constant amino acid sequence. A second consideration concerns the number of inserted amino acids resulting from the linker. It can be desirable to minimize the disruption to the natural proteins by

using as short a linker as possible. However, this can present challenges because it is difficult to overlap the linkers without introducing changes to the OE (or ME) and IE sequences that are detrimental to transposition frequency. The compressed linkers of Fig. 1 are likely to be the shortest possible linkers that employ these binding sequences. As noted above, Linker A has an IE end and an OE end; Linker B has an IE end and an ME end (see Fig. 1).

[0046] If no overlap is provided in the linker (and none is required in this or any disclosed embodiment), a thirty-nine nucleotide long linker would include a complete OE (or ME) sequence, a complete IE sequence, and a single nucleotide therebetween to bring the length of the linker to a multiple of three. These considerations are of no concern when using the invention for manipulations other than gene fusions. Indeed, in the subsequent embodiments, the linkers can be overlapping, abutted or non-overlapping unless another aspect of the method dictates a separation between mismatched transposase-interacting sequences.

[0047] It is instructive to look first at the simplest embodiments of the invention, namely processes for generating libraries of nucleic acid molecules such as plasmids that contain a pair of fused genes that can encode chimeric proteins as shown in Figs. 2 and 3. A pair of distinct protein-encoding (or partial protein-encoding) sequences (Gene A and Gene B) are provided on separate molecules. A goal of these methods for forming fusion proteins is to bring together N-terminal portions of one protein-encoding sequence with C-terminal portions of a second protein-encoding sequence, where the library of resulting fusions includes a random set of junction points. The gene portions can be from unrelated or related genes. This can be accomplished in either of two approaches that share a single underlying principle, although the constructs employed differ slightly, as is discussed below. Also, the construct that contains sequences encoding the C-terminal portion of the fusion can optionally further comprise another polynucleotide fused to the sequence that encodes the C-terminus, where the additional polynucleotide encodes a protein that can be directly selected for. This technique not only selects expressed fusion proteins for screening, but also eliminates translated fusions that result in proteins that are poorly expressed or have low solubility (Maxwell *et al.*, 1999). The successful use of a CAT gene for selecting non-homologous gene fusions has previously been reported. (Sieber *et al.*, 2001).

[0048] In one approach, depicted in Fig. 2, a transposon that confers resistance to selectable marker SMI is transposed into a first gene using Tnp1 to produce an initial insert library. Using Tnp2, the products of that library are transposed into a second gene to form fusion proteins. More particularly, each of a pair of genes A and B is provided on a separate construct

that also comprises an origin of replication and distinct selectable markers, SM2 and SM3, respectively. Gene A contributes its N-terminal portion, and Gene B contributes its C-terminal portion, to the fusion protein. In this approach, the construct that contains Gene A also includes a single Tnp2 transposase-interacting sequence upstream of the gene's promoter.

[0049] The aforementioned Gene A construct is then mixed with a pre-cleaved transposon that includes yet another distinct selectable marker (SM1) flanked at a first end by a first Tnp1-specific transposase-interacting sequence, and at a second end by a transposase-interacting sequence pair having a Tnp2-specific sequence and a second Tnp1-specific sequence, where the Tnp2-specific sequence is between the selective marker and the second Tnp1-specific sequence. In a first transposition reaction between these two nucleic acid molecules in the presence of an amount of Tnp1 effective to catalyze transposition, the transposon inserts into the target construct at random positions, some of which are in Gene A. Reaction products are then transformed into suitable host cells, such as *E. coli* cells, under standard transformation conditions.

Electroporation is a suitable transformation method. The transformed cells are grown under selective pressure (SM1 and SM2) for cells that contain constructs having the Gene A construct and an integrated transposon. The resulting selected cells constitute the initial library of random linker inserts.

[0050] In the presence of an amount of Tnp2 effective to catalyze transposition, DNA from the initial insert library serves in a second transposition reaction as the transposon donor. The initial insert library DNA is mixed with the target Gene B construct and incubated until transposons are cleaved from the initial insert library and then inserted at random locations into the target construct to create the library of random fusion constructs that can encode random fusion proteins having an N-terminal end contributed by Gene A and a C-terminal end contributed by Gene B. The products are transformed into suitable host cells and selected for SM3 and against SM1 and SM2, which indicates the presence of Gene B construct without the Gene A construct or the original transposon of the first reaction. It should be noted that the Tnp2-interacting sequence in half of all inserts in the resulting initial insert library is inverted. Such inserts cannot participate in the second transposition reaction.

[0051] As mentioned above, one can alternatively introduce the transposase-interacting sequences entirely by transposition, instead of providing a single Tnp2-interacting sequence on the Gene A construct. In this second approach, two fusions are created. This second approach also employs a pre-cleaved transposon and separate Gene A and Gene B constructs. Neither of the two constructs contains transposase binding sequences at the start of the method, although

one of the two constructs contains a conditional origin of replication to prevent improper replication of the dual-origin fusion construct that is formed in the method by joining the two starting constructs. Since some fusion constructs can have two origins, improper replication could result if both starting constructs contain non-conditional origins of replication.

[0052] Fig. 3 depicts this second approach. In contrast to the first approach, the orientation of insertion is not important, as the double ends exist on both ends of the transposon. In the first transposition step, parts of genes A and B are transposed to yield a library containing two classes of fusion products, namely (1) the N-terminal end of Gene A fused to the C-terminal end of Gene B and (2) the N-terminal end of Gene B fused to the C-terminal end of Gene A, which can be transcribed and translated to produce fusion RNAs and fusion proteins.

[0053] To carry out the first transposition step in this second approach, a pre-cut transposon encoding selectable marker SM1 flanked by pairs of transposase-interacting sequences as shown and described is combined with a replication competent circular nucleic acid molecule that contains an origin of replication (which can be a conditional origin of replication such as a  $\pi$ -dependent ori) and a selectable marker as well as a gene sequence (Gene A) that can encode a protein. Gene A should include a transcriptional promoter upstream of the gene and a stop codon at the end of the coding sequence. The two are combined in the presence of a first transposase (Tnp1) that preferentially binds to and interacts with the pair of transposase-interacting sequences shown in black under conditions that promote *in vitro* transposition.

[0054] Plasmids in the library of *in vitro* transposition reaction products can be introduced into host cells by electroporation and cells that contain a plasmid having a transposon insertion in the protein-encoding gene of interest are obtained by plating the host cells on a medium that selects for resistance to SM1 and SM2. The selection method ensures that no member of the library has received an insertion in either the origin of replication or the selectable marker provided on the plasmid. A short linker segment remains at each fusion junction. The junction point at which the coding sequence of the first protein ends is randomly determined by this first *in vitro* transposition (strand transfer) reaction. These libraries typically contain a sufficient number of independent survivors ( $>10^5$ ) to encompass all possible insertion sites, as determined by plating of a small dilution of the transformation mixture onto agar plates in the presence of both drugs.

[0055] In a second *in vitro* transposition reaction shown in Fig. 3, DNA from the members of the initial library are mixed *in vitro* under conditions that favor *in vitro* transposition with a second plasmid containing an origin of replication, a third selectable marker SM3, and a second

gene of interest (gene B optionally with promoter) which is intended to be fused with gene A. The origin of replication can be a conditional origin of replication if the origin of replication used in the first step was not a conditional origin. In this second round of *in vitro* transposition, plasmids from the first transposition library and the second plasmid are mixed under suitable conditions with the second transposase (Tnp2) which binds to and interacts with the pair of inverted repeats shown in white. In the resulting *in vitro* transposition, the selectable marker present on the original transposon is lost, as it forms the donor backbone DNA, and the remainder of the library plasmids are the transposable portions transposed into plasmid B. After transformation into a suitable host, those colonies that grow in the presence of SM2 and SM3 contain fusions as shown at the bottom of Fig. 3. The resulting reactions products can be directly transformed into suitable host cells and grown in mixed liquid culture under double drug selection to generate a library of desirable products. Plating of small aliquots from reactions has revealed that the library can contain  $>10^7$  independent gene fusions. Alternatively, the reaction products can be subjected to selection of functional fusion products. These fusion polynucleotides can encode fusion RNAs that can be translated to produce fusion polypeptides (also referred to interchangeably herein as proteins). The total size of an individual chimeric protein can vary from zero to the total length of both proteins plus the linker segment. One can analyze the nucleic acid members of a fusion library thus created using standard cleavage and sizing methods to demonstrate the presence of fusion products.

[0056] This fusion method can be accomplished with no intermediate transfer into a host cell by relying on a single selection after the second transposition process, although this approach is disfavored because of reduced efficiency and, consequently, reduced library size. Because the transposition products in the first stage do not represent a large percentage of the total DNA, it is instead advantageous to purify the initial products so that the ratio of reactants in the second stage can be controlled to promote efficient formation of the desired products.

[0057] It is noted that many microbial selection techniques are known, including without limitation the use of selectable markers and biosynthetic and auxotrophic selections, and any such technique can be employed in the method. Selectable markers are employed in the exemplified embodiments because they are convenient and easily manipulated. A skilled artisan will understand that the selectable marker resistance gene chosen for use at any stage of the methods is not critical and all that is required is that the selection technique be appropriate for the chosen resistance gene. By way of non-limiting example, the marker can confer resistance to kanamycin, chloramphenicol, ampicillin, tetracycline or other drugs.

[0058] When manipulating the nucleic acid and cells as described herein, commercially available *E. coli* K-12 DH5 cells (recA) or JM109 cells are suitable, if no conditional origin of replication is in use. In the latter case, cells that supply the protein required for conditional replication (such as TransforMax EC100D *pir*-116 electrocompetent *E. coli* K12 cells, commercially available from Epicentre, Madison, WI) are suitable. For deletion methods any bacterial strain that supports Tn5 transposition, preferably an *E. coli* strain such as MG1655 (ATCC 47076), can be used.

[0059] The number of actual transpositional protein fusions is limited both because the correct reading frame is maintained in only 1 of 3 events and also because the transposition insert is oriented properly in only 1 of 2 events in the second reaction. Therefore, a maximum of 1 out of 6 (16.7%) of all transpositional fusions can be true protein fusions. Although this limitation on the number of productive fusion products can be ignored when the library is large, as here, it may be otherwise desirable to screen for functionally productive fusions.

[0060] The percentage of functional fusion products is further limited both by stop codons in the linker and by non-essential DNA other than the gene of interest in the construct. The former can be overcome or reduced as needed by modifying the linker sequence as appropriate to eliminate stop codons. While certain such changes could alter or even destroy the transposition reaction efficiency, it is likely that not every change that removed a stop codon would also affect transposition efficiency. The latter can be overcome by careful engineering constructs to ensure that only inserts into the gene of interest can produce productive transposition products that can both replicate and encode proper drug resistance.

[0061] Essentially the same principles apply in the additional methods disclosed below. However, each of the following methods employ as a first step the synaptic complex formation and electroporation method disclosed in incorporated International Patent Application Number WO/00/7343. Briefly, as shown in Fig. 4 a precut transposon containing two pairs of transposase-interacting sequences arranged as disclosed in Fig. 4 is incubated with Tnpl that binds to and interacts with the pair of black transposase-interacting sequences in the absence of magnesium. The initial incubation step forms a synaptic complex poised for a first transposition event in the presence of a target DNA. Upon introducing the synaptic complex into cells by, e.g., electroporation, the synaptic complex transposes into the bacterial chromosome. After introduction into the host cells, those cells that have incorporated the transposon into their chromosome can be selected by screening for the presence of both SM4 and SM5.

[0062] All colonies produced in the first stage of these transposition methods can inducibly encode Tnp2. To induce expression of Tnp2, a colony or set of colonies are inoculated into a selection-free liquid medium containing an inducing agent. The choice of an inducing agent is entirely up to the convenience of the user and has no bearing upon the operation of the invention. A suitable inducing agent would be arabinose, where the arabinose promoter is provided upstream of the transposase-encoding sequence on the transposon. Arabinose is preferred because its promoter is tightly regulated. Upon induction, Tnp2 is synthesized, then binds to and interacts with the white pair of transposase-interacting sequences and a second round of transposition results. At some frequency, the transposition is intrachromosomal and as a result, a portion of the chromosome is deleted and the remaining chromosomal material is recircularized. Accordingly, a somewhat smaller chromosome results. The only evidence of the second transposition event is one residual mismatched pair of end sequences as is shown in Fig. 5.

[0063] One can readily screen for those colonies that contain a true deletion event. After induction of the second transposase and overnight growth to permit *in vivo* transposition, the cells are diluted and plated without drug selection (e.g., on TYE plates). Then, individual colonies are replica plated either on permissive medium or on medium containing SM4 or SM5. As shown in Fig. 6, all the colonies will grow on the permissive medium, assuming that none has suffered a deletion in an essential gene. However, only those colonies that are also sensitive to both SM4 and SM5 are in fact deletion mutants.

[0064] A similar approach is utilized to permit the excised portion of the chromosome to be replicated and cloned. In this embodiment of the invention, shown in Fig. 7, a simple addition of an origin of replication in the portion of the transposon that is not between the inner (white) pair of transposase-interacting sequences accomplishes this goal. It is important to take due care to ensure that the origin of replication employed is not detrimental or lethal to the host cell. The transposon is introduced into the chromosome as in Fig. 4. Likewise, Tnp2 is induced in precisely the same way. Upon intrachromosomal transposition, however, the excised portion, shown at the bottom right of Fig. 7, contains an origin of replication and can be maintained as an extrachromosomal plasmid.

[0065] Finally, relying upon precisely the same principles as the preceding two embodiments, the next embodiment facilitates insertion of a DNA of interest into the chromosome in place of the deleted portion. In this embodiment, shown in Fig. 8, the insert DNA is provided on the transposon outside of the inner (white) pair of transposase-interacting

sequences but not on the segment of the transposon that includes SM5. This portion of the transposon remains in the chromosome after the second round of transposition. As is shown in Fig. 8, SM5 is lost with the deleted portion of the chromosome while the desired insert is inserted in place of the deleted portion.

[0066] In any of the disclosed embodiments, the products of a first double transposition method can, in fact, act as substrates for a second round of double transposition thereby further increasing the variety of the products. In particular, the chromosomal deletion and gene fusion embodiments are well suited to practicing multiple iterations of the method. By coupling the ability to produce a wide variety of fusion proteins with a known ability to screen for those proteins having desired structures or functions makes the disclosed method of great interest to the pharmaceutical industry, among others. At least the cloning and deletion methods can also be practiced entirely *in vitro* on non-chromosomal targets (e.g., plasmids).

[0067] To obtain products of the methods shown in Figs. 7, one can use a regulated origin of replication on the resulting plasmid to provide additional level of selection. To obtain products of the methods shown in Fig. 8, one can merely replica plate the induced cells and select or screen those that have lost or retained the selectable marker or markers of interest, as appropriate for each method.

[0068] Direct analysis of the random, non-lethal chromosomal deletion products can reveal the largest deletion that can support growth in particular medium. Moreover, because these deletion products leave behind only a small DNA segment and no selectable markers or transposase coding sequences, the process can be performed recursively, such that the deletion product is used in a subsequent round of insertion or deletion. By repeating this process several times, it is possible to reduce the size of the genome, even to the point of producing a bacterial strain that contains a chromosome having only those genes essential for reproduction.

[0069] Additionally, a key requirement in high throughput DNA sequencing is to have a technology in place that will allow a defined primer binding site to be moved next to the region to be sequenced. A single transposition insertion event "randomly" places 2 primer binding sites within a target DNA molecule. In contrast, the double transposition deletion formation system of the invention can advantageously generate multiple constructs from a single transposon insert thereby "moving" a single primer binding site to a plurality of target DNA locations. An embodiment of this methodology is described in the following paragraphs.

[0070] A transposon is constructed with distinct pairs of nested transpose-interacting sequences in inverted orientation relative to each other (as described elsewhere herein). For

example, if the outer pair are IE sequences and the inner pair are ME sequences, the IE sequences with an appropriate Tnp1 are used to generate initial insertions and the ME sequences with an appropriate Tnp2 are used to generate adjoining deletions as described. Appropriately placed inside the element are suitable selective markers and primer binding sites.

[0071] The transposon is inserted at random using the IE specific transposase *in vitro* into a target cloned DNA, such as a BAC clone. The transposon inserted target clone is transformed into cells and independent transformants are isolated and grown up. DNA sequencing is performed using two primers off of the two ends of the transposon inserts.

[0072] From each insert, or from selected inserts, multiple adjacent deletions are generated using Tnp2. This can be accomplished either *in vivo* (requiring prior transformation and then *in vivo* synthesis of transposase) or *in vitro* followed by transformation. A suitable primer is then employed to sequence across the deletion end points into adjacent sequences. Each insert and deletion can also be physically mapped using appropriate restriction digests if desired.

[0073] Although the methods of the invention are exemplified using transposable polynucleotides having two transposase-interacting sequence pairs that interact with two transposases, this is not intended to be a limitation on the invention, since one can engineer any greater number of transposase-interacting sequences into a transposable polynucleotide, as long as a transposase having a distinct activity can be used in combination with each such sequence pair.

#### Example 1

##### Restoration of chloramphenicol acetyl transferase (CAT) from N- and C-terminally-truncated genes

[0074] Plasmid Vectors

[0075] Plasmids were maintained in *E. coli* K12 (DH5 $\alpha$ ), except plasmids that contained a  $\pi$ -dependent origin of replication which were maintained in EC100 pir-116 cells. Plasmids were purified using a quiafilter midi kit (Quiagen). Pre-cleaved transposons were released from the appropriate purified plasmids by cleavage with a restriction enzyme and were then purified by electrophoresis and gel extraction using the quiaquick gel purification kit (quiagen).

[0076] The plasmid vectors for the first approach to gene fusion were as shown in Fig. 2. A first plasmid containing a kanamycin resistance gene flanked by one IE and one IE/OE linker end (Linker A) was constructed using standard techniques from starting plasmid pGT4 (SEQ ID NO:8) which encodes a kanamycin resistance gene flanked by IE sequences. This plasmid was the source of pre-cleaved transposon, which was released from the resulting plasmid by cleavage

with PshA1. A second plasmid contained a C-terminally truncated CAT gene fragment encoding the first 171 of 219 amino acids and an upstream promoter next to a transposon-interacting sequence specific for ME, as shown at the top of Fig. 2. The source of the CAT gene fragment was pACYC184 (New England Biolabs). The CAT gene was cloned into a vector between ME sequences. The vector also included an origin of replication and a gene conferring resistance to ampicillin. The 3'-end of the gene and the adjacent ME sequence were deleted from the plasmid by cleavage with AfII and subsequent religation. This plasmid was the target in the first transposition reaction. A comparable approach was taken to generating a third plasmid that contained an N-terminally truncated CAT gene fragment encoding all but the first 38 amino acids. In this case a short PvII fragment was excised from pACYC184 and then re-ligated. The third plasmid also included an origin of replication and a gene conferring resistance to tetracycline. This plasmid was the target in the second transposition reaction.

[0077] The plasmid vectors for the second approach to gene fusion were as shown in Fig. 3. The source of pre-cleaved transposon substrate for the first transposition reaction was like that of the first approach, except that Linker A (or, in an alternative embodiment, Linker B) was provided in appropriate orientation at both ends of the transposon that confers resistance to kanamycin. The second plasmid contained a full length CAT gene (Gene A), an origin of replication and a gene that confers ampicillin resistance. The third plasmid contained an N-truncated CAT gene (Gene B) and a pi-dependent origin of replication. It will be appreciated that either the second or the third plasmid can be provided with the pi-dependent origin, which is advantageously used in maintaining members of interest in a library. Similarly, it is understood that the genes provided on the plasmids can be full length or partial length, depending upon the nature of the genes and the goal of the particular transposition process.

#### Gene fusion protocol

[0078] In all transposition reactions, DNAs were combined with the appropriate Tnp in transposition reaction buffer (0.1 M potassium glutamate, 25 mM Tris acetate, pH 7.5, 10 mM Mg<sup>2+</sup> acetate, 50 µg/ml bovine serum albumin, 0.5 mM β-mercaptoethanol, 2 mM spermidine, 100 µg/ml tRNA; final concentrations) in a total volume of 20µl. Tnp1 and Tnp2 were purified using the IMPACT system (New England Biolabs) as described previously (Bhasin *et al.*, 1999).

[0079] In the first transposition reaction, pre-cleaved transposon and target plasmid were added to a final concentration of 20nM each. Tnp1 was added to a final concentration of 200nM and the reactions were incubated at 37°C for 3 hours. Products were then subjected to treatment

with SDS (0.5%) and heat (68°C for 5 minutes) to remove Tnp from transposition products and to increase transformation efficiency (data not shown). Treated reaction products were then dialyzed against water and transformed into electrocompetent *E. coli* K12 cells (DH5α; efficiency = 8.0X10<sup>8</sup> cfu/μl plasmid). Cells were grown at 37°C in 1 ml of Luria-Bertani Medium (LB, Difco) to allow expression of antibiotic resistance genes. The culture was then used to inoculate 50ml of LB containing appropriate antibiotics and grown at 37°C for 16 hours. A small aliquot of culture following the initial 1 hour growth was removed and plated onto agar media to estimate library size (>10<sup>5</sup>).

[0080] In the second transposition reaction, purified plasmid DNA from the insert library and the second target plasmid were added to a concentration of 80nM each. Tnp2 was then added to a concentration of 400nM. Reactions were allowed to incubate at 37°C for 1 hour and then treated with SDS and heat, dialyzed, and transformed as above.

[0081] In the second approach, the initial insert library was generated by mixing pre-cleaved transposon having Linker A at both of its ends with a target plasmid that contained sequences for the N-terminal truncation of the CAT gene in the presence of Tnp1. The library was transformed into *E. coli* host cells which were outgrown in 1 ml of LB without drugs for 1 hour and subsequently inoculated into 50ml LB and allowed to grow overnight in the presence of double antibiotic selection. The insert library was then purified. In the second reaction, plasmid DNA from the insert library was then mixed with the second plasmid that contained that contained sequences for the C-terminal truncation of the CAT gene. Tnp2 was added to cleave N-terminal fragments from the CAT gene in the insert library and insert them randomly into the second target plasmid. The reaction products were then transformed into *E. coli* host cells and plated on LB agar containing 20μg/ml chloramphenicol to select for functional CAT fusions.

[0082] Although successful, Linker A was inefficient, so Linker B was tested in the same way. Due to the order in which the two plasmids were used, the linker orientation for fused genes was in the opposite direction (ME to IE) to those created by method one.

[0083] Gel electrophoresis confirmed that the plasmid fusion libraries of the second approach contained one major band, indicating that the double drug selection efficiently chose the desired products from the second transposition reaction without prior purification. Fusion plasmids were shown to contain a unique EcoRV site from the first target plasmid and a unique ScaI site from the second target plasmid. Double digestion of a single fusion plasmid yielded two linear DNA bands having a combined size equal to that of the fused plasmids (not shown). However, as expected, a smear of DNA bands of varying size was observed when the library was

double-digested with EcoRV and ScaI because the size of the two bands in each fusion plasmid is different, given that the two plasmids fuse at random positions in each individual product.

[0084] To confirm the presence of the transposon linker sequence and to determine the amino acid sequence of the functional fusions, plasmid DNA from a few chloramphenicol resistant colonies obtained in each approach were purified and sequenced. No matter which approach was taken, functional fusions were obtained. Eight functional fusions (obtained from both approaches) were analyzed. None was a ‘perfect’ fusion (showing no loss or duplication of amino acids). Seven of the eight included a duplicated CAT fragment, ranging in length from 5 amino acids to 54 amino acids in length. One fusion was formed by a 9 amino acid deletion.

[0085] Structural information for the CAT protein used in this study, which originates from transposon Tn9, does not exist. However, X-ray crystallographic structure information is available for the related type III Chloramphenicol Acetyltransferase (CATIII) (Leslie, 1990). A BLAST alignment of the primary amino acid sequences of CAT and CATIII revealed that the two proteins have 46% amino acid identity with only a single gap. This alignment was used to map the location of linker insertion in functional CAT fusions to the CATIII structure. This analysis predicts that, of the sixteen linker-CAT junctions, 15 are inserted in disordered regions, while the remaining junction is predicted to be inserted into a  $\beta$ -sheet one amino acid from the terminus. Although these results come from only one protein fusion example, the analysis strongly suggests that functional fusions are formed when the linker sequence does not insert into a location that interrupts either  $\alpha$ -helices or  $\beta$ -sheets.

## Example 2

### Deletions and Identification of Essential Genes

[0086] Media. For all experiments we used Luria broth liquid or plates media adjusted with antibiotics if needed as follow: chloramphenicol 20 mg/L, kanamicin 40 mg/L, ampicillin 100 mg/L.

[0087] Bacterial strains. E.coli K12 strain MG1655 with a known sequence of chromosomal DNA (Blattner et al. 1997) was used for making deletions. E.coli K12 strain DH5 $\alpha$  (Sambrook et al. 1989) was used for DNA manipulations.

[0088] Fig. 9 depicts the transposons used in this Example. Transposon TnDEL7 was designed for inducing deletion of dispensable genetic material and does not have an origin of replication, thereby allowing the immediate isolation of deletions after induction. Transposon

TnDEL8 contains a conditional (regulated) origin of replication and was used to maintain deleted genetic material in the cell on a plasmid until elimination is triggered, as described below. The selectable markers and other attributes of each transposon are as shown in Fig. 9. Both transposons Tn5DEL7 and Tn5DEL8 were cleaved from donor vectors pGT7 (SEQ ID NO:9) and pGT8 (SEQ ID NO:10), respectively, by restriction enzyme digestion using PshAI and were extracted from agarose gel using QIAquick Gel Extraction Kit (Qiagen).

[0089] Both transposons contain nested pairs of transposase-interacting sequences. The outer pair of sequences are Tn5 IE ends that can participate in forming synaptic complexes in the presence of Tnp1. The inner pair of sequences are ME artificial ends that can participate in transposition reactions in the presence of Tnp2. This efficient combination is preferred, as high transposition efficiency is important for effective screening without selection.

[0090] In the following trials, transposome synaptic complexes were introduced into cells by electroporation. Km<sup>R</sup> and Cm<sup>R</sup> colonies were selected. For Tn5DEL7, 50 to 80 percent of transposition events were recovered, with a quarter having deletions of desired orientation. For Tn5DEL8 the percent of successful deletion events exceeds 95%, presumably due to negative consequences of having two active origins of replication on the chromosome.

[0091] Depending upon the task, single colonies, or a pooled collection of about 50 colonies, were used to start liquid cultures. At an early exponential growth state, arabinose was added to induce a second transposition step. In the case of TnDEL8, chloramphenicol and IPTG were also added. IPTG maintains plasmids formed during deletion induction and to depress growth of cells in which the transposon-encoded ori did not excise from the chromosome during induction. This creates a selection against such cells. After a few hours, cells were sub cultured in the same media with 100-fold dilution and left shaking overnight. Then, cells were diluted and plated to yield single colonies for replica plating. Km<sup>S</sup>, Cm<sup>S</sup> cells were picked in the case of TnDEL7 and Km<sup>S</sup> cells in the case of TnDEL8. In the latter case, plasmid DNA was prepared for analysis.

[0092] First, deletion size distribution was examined in the lactose operon region. This region was chosen for study since it is known to allow large deletions of at least 100Kb (Bachmann, 1996). Tn5DEL7 was delivered as a synaptic complex into *E. coli* MG1655 cells that were then plated on Lactose-McConkey's medium with antibiotic selection. *E. coli* MG1655 with the deletion transposon inserted into the lac operon was isolated. A few white colonies were selected from thousands of red colonies. One colony was chosen for sequencing of the chromosomal DNA around the transposition site. The insert was located at 362,522-31bp of the

*E.coli* map. A single colony having the *lac* phenotype was selected and DNA sequencing was performed to determine the location of the insertion.

[0093] Then deletion formation was induced by induction of Tnp2 with arabinose and the deletions were collected by replica plating. Screening for deletions in the case of Tn5DEL7 was made by replica plating colonies on LB agar, LB-kanamycin (20mg/L), and LB-chloramphenicol (20mg/L). Cells sensitive to kanamycin were considered to have undergone transposition, and cells sensitive to both drugs were considered to have a deletion of adjacent DNA and the "right" portion of the transposon with only a small transposon linker being left on the chromosome.

[0094] A total of 9 independent deletions were analyzed by sequencing performed on ABI PRISM model 377, either directly from chromosomal DNA or by an inverse PCR technique. For *lacZ* insertion, direct chromosomal DNA sequencing was performed using primer FWD2: 5' CAGATCTCATGCAAGCTTGAGCTC 3' (SEQ ID NO:5), complementary to the transposon linker. For sequencing deletions produced from this insertion, inverse PCR was performed with primers GGTCTGCTTCTGACAAACTCGGGC (SEQ ID NO:6), and ACGCGAAATACGGGCAGACATGGCC (SEQ ID NO:7). after digesting the chromosomal DNA with FspI and ligating. PCR products were purified from an agarose gel and sequenced with the standard Big Dye protocol. Of these nine deletions, two events were insertions within the transposon that resulted in loss of chloramphenicol resistance. The remaining seven were desirable events that deleted various lengths of chromosomal DNA. The summary of the sequencing analysis is shown in Fig. 10. The deletions vary in size from 4 to 23Kb, with most deletions being about 20Kb. The average deletion size is both surprising and encouraging. First, transposon attacking its own DNA is expected to find the target at a distance just beyond the persistence length, which is expected to be about 100 to 200 base pairs in the cell. Secondly, with deletions of this size, it is feasible to reach saturation (no further deletions without deleting essential genes), under conditions that permit repetitive application of the technology in a reasonable time frame.

#### Recursive deletion formation

[0095] Advantageously, all transposon components, apart from a short linker (64bp), are lost after a deletion is generated using this system. The loss of the transposase gene ensures that the chromosome remains stable (in terms of subsequent transposition transposition). The loss of all selectable markers provides an opportunity to iteratively repeat the process.

[0096] Accordingly, 20 consecutive rounds of deletions were induced in MG1655 cells as follows. The protocol was applied repeatedly to mixtures of at least 10 independent deletions, thereby avoiding isolating deleterious deletions that result in delayed cell growth in a mixed culture. After 20 rounds, ten final strains having growth rates at least equal to that of the parental strain were obtained. After analyzing the new strains for the loss of metabolic activities, it was concluded that the strains are in part interrelated and in part independent.

[0097] To analyze the diversity and average size of deletions in these isolates, four isolates were digested with NotI and the digests were separated by pulse field electrophoresis using a CHEF-DR II BioRad system, according to the protocol of Heath et al., 1992, with minor technical modifications. The pattern for MG1655 matches that previously reported, and the identification letter for each fragment is shown at right. For each deletion strain, the pattern is altered and indicates a loss of DNA. The size of each new band in the four deletion strains was estimated by comparison to DNA markers. In some cases, bands seen in digested MG1655 DNA are absent, replaced in the isolates by shorter or longer fragments. The total amount of DNA deleted for four of the isolated strains was calculated as 250, 262, 100, and 247kb. Thus an average deletion size per round is 11 kb, which agrees well with results obtained for the lactose operon region. It is assumed that the size of some deletions is limited by the presence of nearby essential genes. This suggests that large pieces of DNA can frequently be deleted during the first random chromosomal insertions. In other words, *E.coli* has significant excess DNA for growth in rich media. Second, 20 rounds are insufficient to saturate the chromosome with deletions.

#### Coupled chromosomal deletion / plasmid formation system.

[0098] The obvious limitation of the deletion formation system, described above is the inability to introduce deletions that involve essential genes. It would be advantageous to save deleted DNA conditionally in the same cell and to attempt to eliminate it later. Transposon Tn5DEL7 was, therefore, modified to include a conditional origin of replication, namely the origin from pAM34 (ATCC 77185). The conditional origin was easily controlled by IPTG, maintained moderate copy number, was lost relatively quickly under nonpermissive conditions, and was competent to accommodate large inserts. Importantly, the origin characteristically shows very low background on selective media with selection for the presence of drug resistance or essential gene. Only one or two colonies can arise on a plate lacking IPTG.

[0099] The protocol is as follows. Synaptic complexes were formed by mixing Tnp1 with precut Tn5DEL8 in binding buffer. Transposome complexes were assembled by incubation of

precut transposon with Tnp1, 20 mM Tris Ac, 100 mM K Glutamate for 1h at 37 °C. Molar ratio was attempted to be 1:5, DNA : protein, with DNA concentration of 0.1ug/ul.

[00100] The complexes were introduced into electrocompetent cells at standard recommended conditions (2.5Kv, 5mS). Km<sup>R</sup>, Cm<sup>R</sup> colonies were selected in the first transposition selection. Individual colonies were inoculated into liquid media. At this point 0.4% arabinose and 1 mM IPTG were added and the cultures were aerated with shaking in 4ml volume in 20ml tubes at 37°C. Arabinose induces synthesis of Tnp2 and, hence, subsequent transposition events. IPTG supports maintenance of the plasmids containing DNA deleted from the chromosome. The cells of interest were kanamycin sensitive and chloramphenicol resistant, since replication of excised DNA circles was supported by keeping 1mM IPTG in the media. Plasmids were detected in more than 90% of cases, indicating strong selection against active extra origin of replication on the chromosome.

[00101] Table 1 shows the final products obtained for 15 independent insertions of Tn5DEL8. The largest plasmid for each case was analyzed by sequencing to define the initial insertion position and the extent of deletion. To determine whether the deleted DNA in each plasmid was essential, cells were streaked on rich media with no IPTG. In the absence of IPTG, the extrachromosomal plasmid, which containing the excised sequences, is lost. In some cases individual colonies did not form, indicating a dependence on the plasmid. The simplest cases is isolate 12.17. Only gene dnaK is deleted; it is known to be essential. In isolate 14.17, one of two genes (rpoZ and gmk) is essential. For insertion #4 three plasmids were analyzed to define DNA carrying essential genes. The difference in size among the plasmids is minimal, but the difference in consequences of elimination of plasmids is dramatic. Plasmids 4.6 and 4.9 differ by two genes, glyS and glyQ. Although strain 4.6 can grow without IPTG, strain 4.9 cannot. Thus, at least one of the genes excised in isolate 4.9 is essential. With the exception of these genes, all other ORFs (complete or interrupted) can be considered non-essential (at least for growth in rich media).

[00102] For both strategies described here, deletions were obtained without direct selection as would be required for, e.g., SucB or GalK. That would require transposons to be larger and more importantly, would restrict the system to use with *E.coli*.

[00103] This Example demonstrates the proof of principle for creating a list of essential and nonessential bacterial genes. It is, accordingly, now technically possible to create a representative library of deletions/complementary plasmids covering a genome multiple times. Then systematic sequencing of deletion borders, in combination with tests of survival after

eliminating the complementary plasmids can provide an extensive (ideally complete) set of genes that can be eliminated without detrimental consequences to the cell.

[00104] The present invention is not intended to be limited to the foregoing, but rather encompasses all such variations and modifications apparent to the skilled artisan that fall within the scope of the appended claims.

TABLE 1

Insert #	$\Delta$ #	Insert End	$\Delta$ End	Size	Partial ORFs	Complete ORFs	Grow w/o IPTG
1	1.3	1073109	1094823	21714	b1028;b1012	b1018;purA;puIP; b1016-1018;ycdB;phoH; b1021-1025;tra5_3;b1027	Y
2	2.8	2776980	2770104	6826	b2638;b2647	b2639;b2646	Y
3	3.1	4407731	4403605	4126	yifl	yifB;yacBYifH	Y
4	4.2	3735047	3725227	9821	yiaB	xyB;xyA;xyF;xy/G;xy/H;xy/R;b3570	Y
4	4.6	3735047	3723521	11526	none	yiaH;yiaA;yiaB;xyA;xyF;xy/G;xy/H;xy/R;b3570	Y
4	4.9	3735047	3719490	15557	y5b	glyS;glyQ;yiaH;yaA;yiaB;xyA;xyF;xy/G;xy/H;xy/R;b3570	N
5	5.1	1224154	1210042	14112	mcra;minD	b1160-b1173;minE	Y
6	6.2	3045069	3042244	2825	bgfA;gcvP	ygfF	Y
7	7.1	2779299	2767042	12227	b2633;b2647	b2634;b2646	Y
8	8.4	3882592	3882500	92	none	none	Y
9	9.7	4589240	4595093	5853	nddB;ter	yifN;yifM;b4356	Y
10	10.3	4141018	4117696	23322	menA;frwC	hsU;hsIV;ftsN;cytR;priA;rpmE;yifX;metU;metB;metL;metF;katG;yifE;yifF;gidA;talC;ptsA;yifJ	N
11	11.3	2793999	2795902	2803	gabP;b2668	ygaE;b2665-b2667	Y
12	12.17	14166	12088	2078	none	dnaK	N
13	13.19	4542104	4546611	4507	fimC;fimH	fimD;fimF;fimG	Y
14	14.17	3821524	3818998	2526	spot	rpoZ;gmk	N
15	15.17	1570886	1544492	26194	yddC;yddG	fdnG;fdnH;fdnI;b1477;psvV;b1481;psmC;b1483-1491;xasA;gadB	Y

## CLAIMS

WE CLAIM:

1. A polynucleotide comprising distinct first and second transposase-interacting inverted repeat sequence pairs, each sequence pair having a specificity for binding to and interacting with a distinct transposase enzyme, members of the first sequence pair flanking members of the second sequence pair.
2. A polynucleotide as claimed in Claim 1 wherein the first or second transposase-interacting inverted repeat sequence pair comprises Tn5 wild-type inside end sequences.
3. A polynucleotide as claimed in Claim 1 wherein the first or second transposase-interacting inverted repeat sequence pair comprises Tn5 mosaic end sequences.
4. A polynucleotide as claimed in Claim 1 wherein the first or second transposase-interacting inverted repeat sequence pair comprises Tn5 wild-type outside end sequences.
5. A polynucleotide as claimed in Claim 1 further comprising between the members of the second inverted repeat sequence pair a first sequence for conferring selectability upon a host cell.
6. A polynucleotide as claimed in Claim 5 further comprising, between the members of the second inverted repeat sequence pair, a polynucleotide that encodes a transposase that specifically binds to and interacts with the second sequence pair and still further comprising, between a first adjacent pair of distinct inverted repeat sequences, a second sequence for conferring selectability upon a host cell.

7. A polynucleotide as claimed in Claim 6 further comprising an origin of replication between the first adjacent pair of distinct inverted repeat sequences.

8. A polynucleotide as claimed in Claim 6 further comprising a preselected polynucleotide sequence insert between a second adjacent pair of distinct inverted repeat sequences.

9. A method for producing a gene fusion library, the method comprising the steps of: mixing together, under conditions suitable for *in vitro* transposition, (1) copies of a transposable polynucleotide that comprises distinct first and second transposase-interacting inverted repeat sequence pairs, members of the first sequence pair flanking members of the second sequence pair, the members of the second sequence pair flanking a first sequence for conferring selectability upon a host cell, (2) copies of a first target nucleic acid molecule that comprises a first polypeptide-encoding gene sequence, a first origin of replication and a second sequence for conferring selectability upon a host cell, and (3) a first transposase that binds to and interacts with the first sequence pair, to produce first transposition products; introducing the first transposition products into host cells; selecting host cells that comprise first transposition products wherein the first polypeptide-encoding gene sequence is disrupted by the transposable polynucleotide; mixing (1) the transposition products from the selected host cells and (2) a second target nucleic acid molecule that comprises a second polypeptide-encoding gene sequence, a second origin of replication and a third sequence for conferring selectability upon a host cell, and (3) a second transposase that binds to and interacts with the second sequence pair, to produce second transposition products; introducing the second transposition products into host cells; selecting host cells that comprise second transposition products comprising fusions between the first and the second polypeptide-encoding gene sequences.

10. A method as claimed in Claim 9 wherein the first origin of replication is a conditional origin of replication.

11. A method as claimed in Claim 9 wherein the second origin of replication is a conditional origin of replication.

12. A method for deleting a portion of a chromosome of a host cell, the method comprising the steps of:

introducing into the host cell a synaptic complex comprising (1) a first transposase and (2) a transposable polynucleotide that comprises distinct first and second transposase-interacting inverted repeat sequence pairs, members of the first sequence pair flanking members of the second sequence pair, further comprising between the members of the second inverted repeat sequence pair a first sequence for conferring selectability upon a host cell and a polynucleotide that encodes a second transposase that specifically binds to and interacts with the second sequence pair and still further comprising, between a first adjacent pair of distinct inverted repeat sequences, a second sequence for conferring selectability upon a host cell;

selecting host cells in which the transposable polynucleotide has integrated into the chromosome;

inducing expression of the second transposase in the selected host cells; and screening to isolate cells in which a portion of the chromosome is deleted.

13. A method for cloning a portion of a chromosome of a host cell, the method comprising the steps of:

introducing into the host cell a synaptic complex comprising (1) a first transposase and (2) a transposable polynucleotide that comprises distinct first and second transposase-interacting inverted repeat sequence pairs, members of the first sequence pair flanking members of the second sequence pair, further comprising between the members of the second inverted repeat sequence pair a first sequence for conferring selectability upon a host cell and a polynucleotide that encodes a second transposase that specifically binds to and interacts with the second sequence pair and still further comprising, between a first adjacent pair of distinct inverted repeat sequences, a second sequence for conferring selectability upon a host cell and an origin of replication;

selecting host cells in which the transposable polynucleotide has integrated into the chromosome;

inducing expression of the second transposase in the selected host cells; and

screening to isolate cells in which a portion of the chromosome is cloned onto a self-replicating nucleic acid molecule that comprises the origin of replication.

14. A method for inserting a preselected polynucleotide sequence insert into a chromosome of a host cell, the method comprising the steps of:

introducing into the host cell a synaptic complex comprising (1) a first transposase and (2) a transposable polynucleotide that comprises distinct first and second transposase-interacting inverted repeat sequence pairs, members of the first sequence pair flanking members of the second sequence pair and comprising between the members of the second inverted repeat sequence pair a first sequence for conferring selectability upon a host cell and a polynucleotide that encodes a second transposase that specifically binds to and interacts with the second sequence pair and further comprising, between a first adjacent pair of distinct inverted repeat sequences, a second sequence for conferring selectability upon a host cell and further comprising, between a second adjacent pair of distinct inverted repeat sequences, the preselected polynucleotide sequence insert;

selecting host cells in which the transposable polynucleotide has integrated into the chromosome;

inducing expression of the second transposase in the selected host cells; and

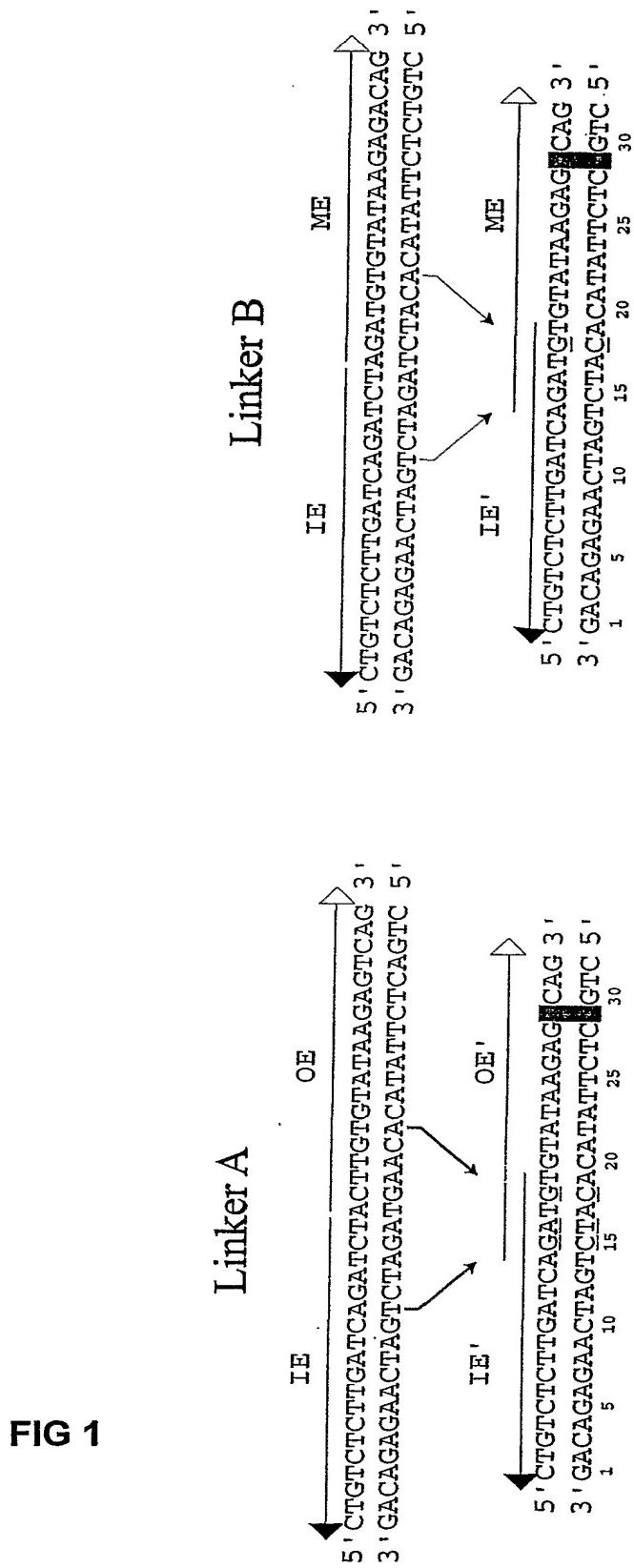
screening to isolate cells in which the preselected polynucleotide sequence insert is inserted into the chromosome and the second sequence for conferring selectability upon a host cell is lost from the chromosome.

15. A polynucleotide fusion product between portions of first and second genes that encodes a polypeptide fusion, wherein the portions of the first and second genes are linked by a pair of non-identical transposase-interacting sequences.

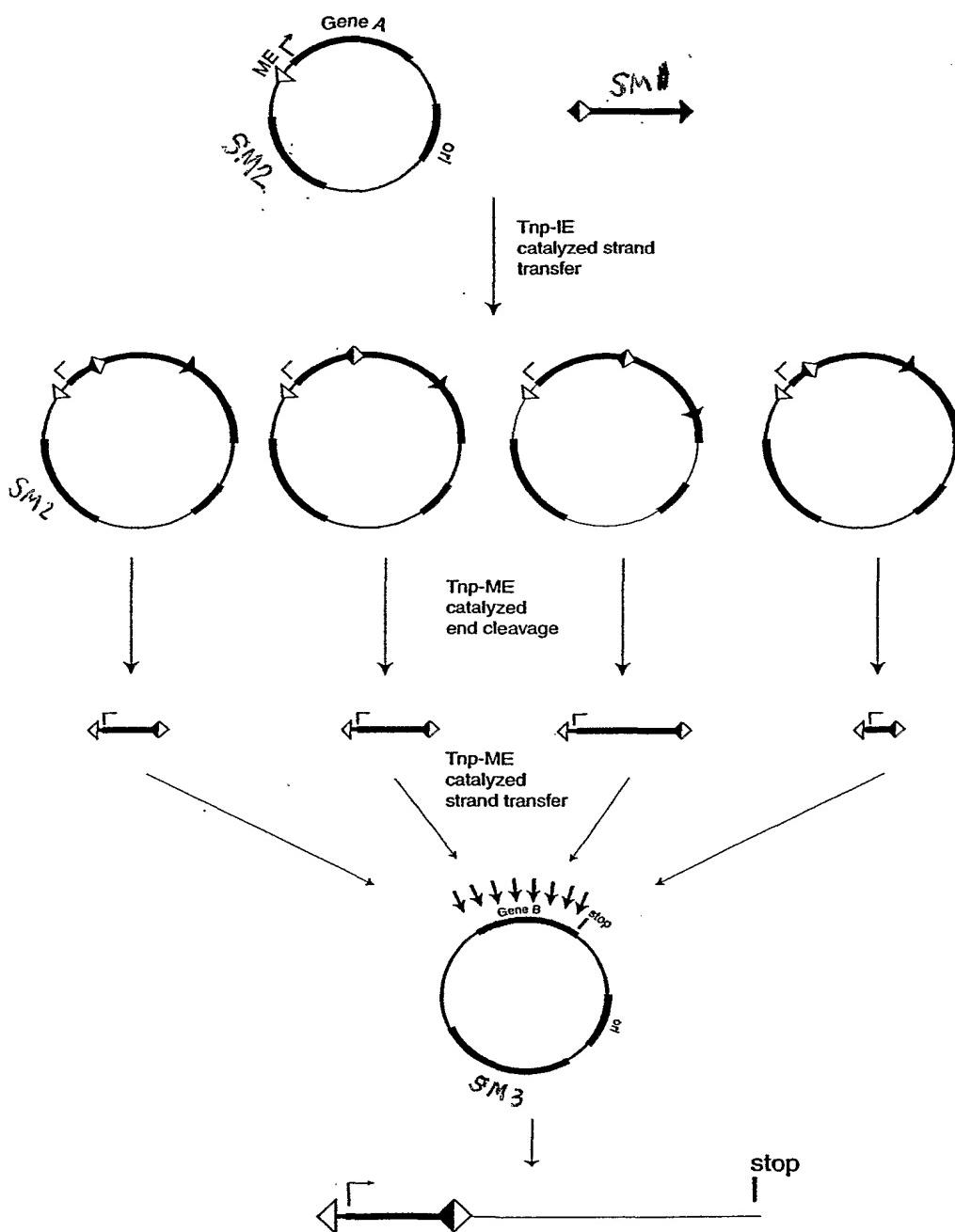
16. A chromosome lacking at a site a portion of nucleic acid and comprising at the site a pair of non-identical transposase-interacting sequences.

17. A chromosome containing at a site an inserted portion of nucleic acid flanked by a pair of non-identical transposase-interacting sequences.

18. A self-replicating nucleic acid molecule comprising a pair of non-identical transposase-interacting sequences flanking a portion of a chromosome and an origin of replication.



2/10

**FIG 2**

3/10

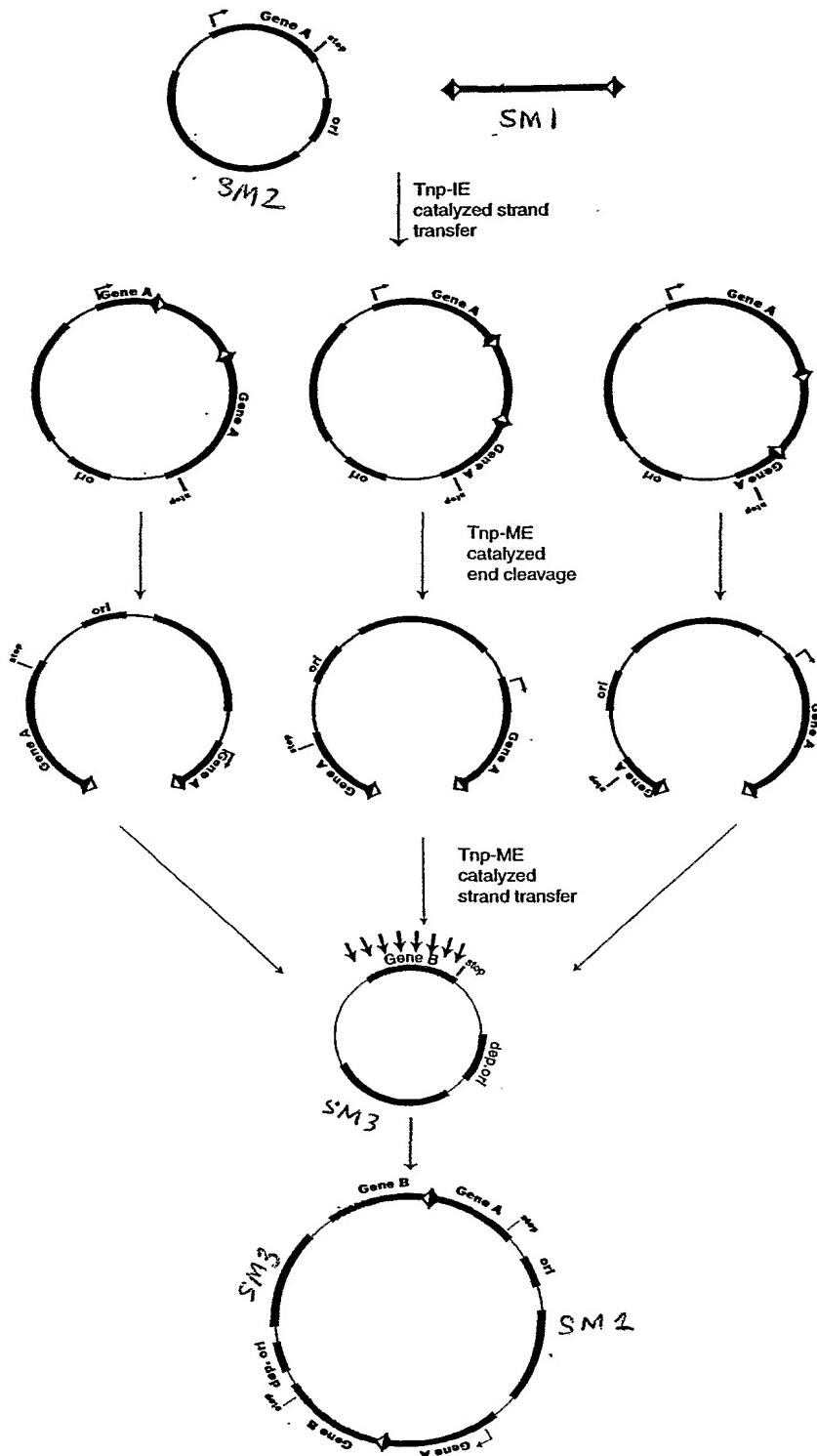
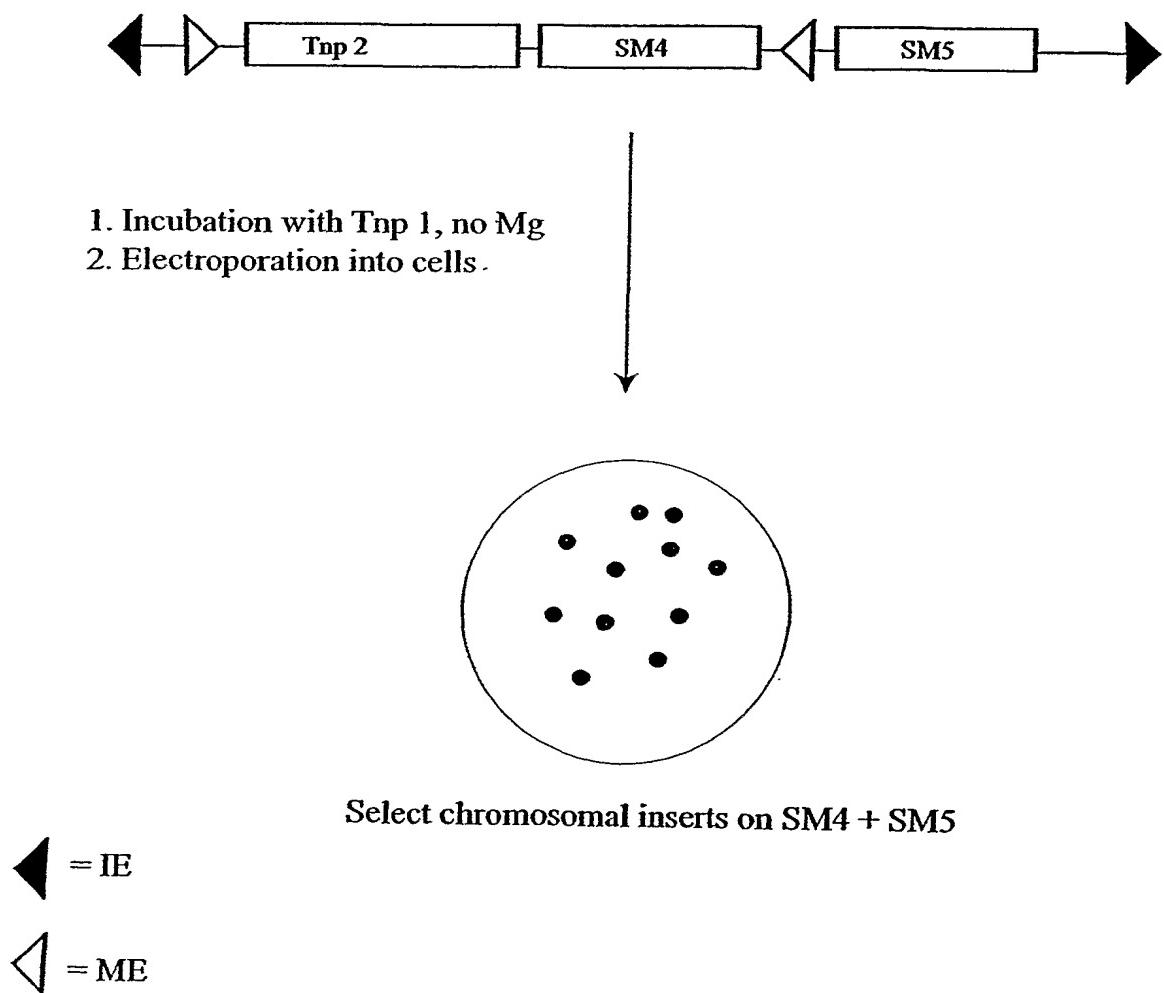
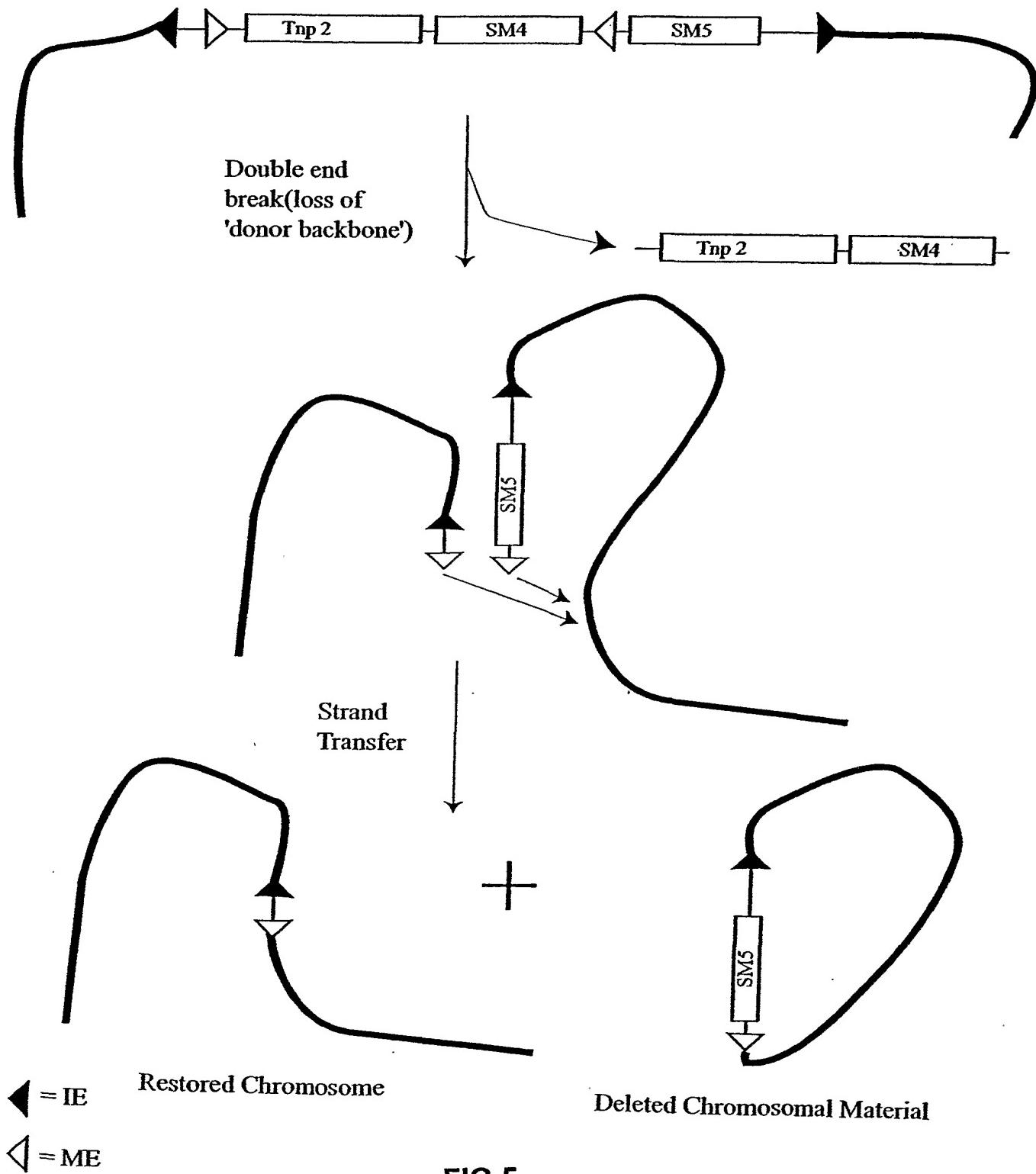


FIG 3

4/10

**Chromosomal insertion of transposon via co-delivery.****FIG 4**

**In vivo transposition.****FIG 5**

6/10

## Screen for chromosomal deletions.

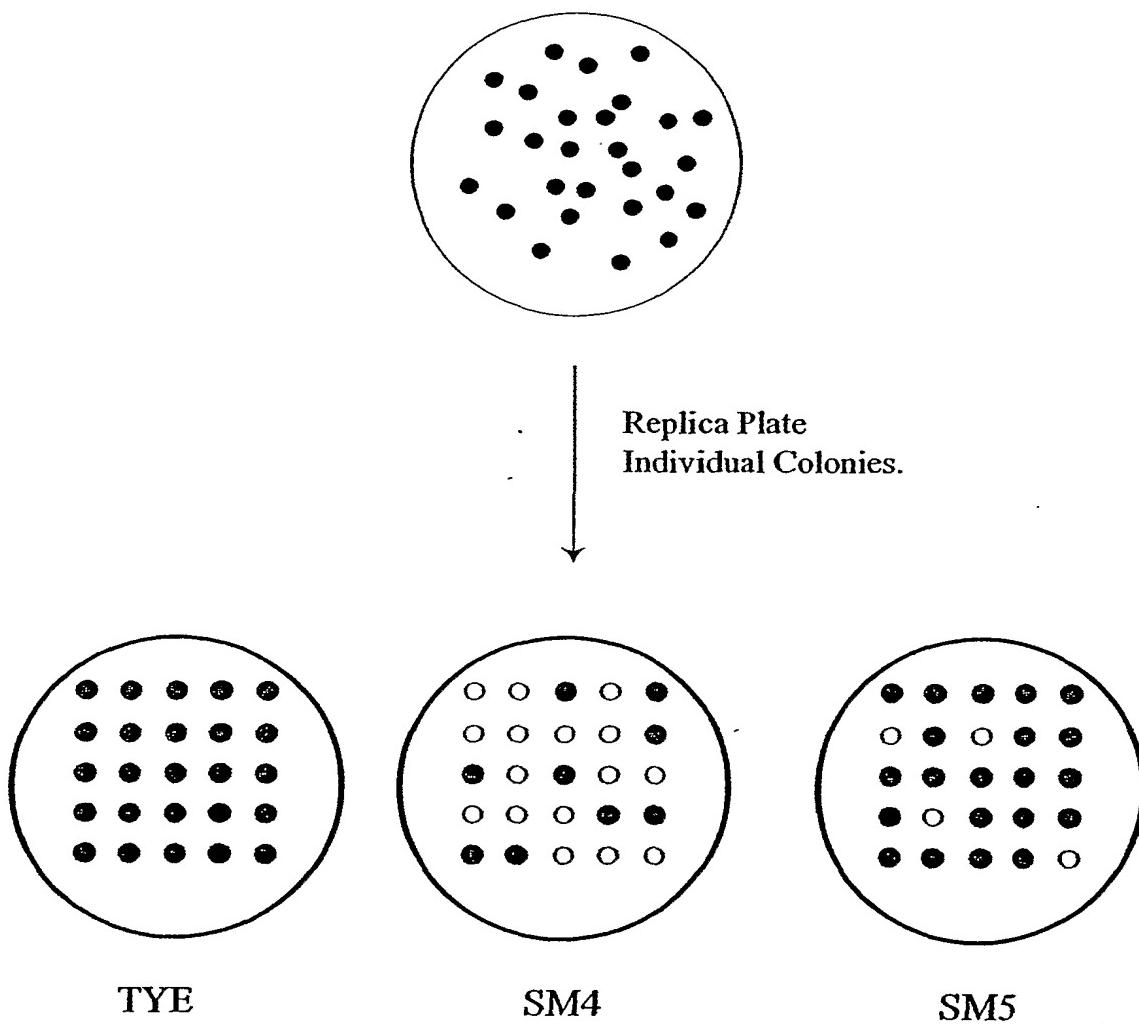
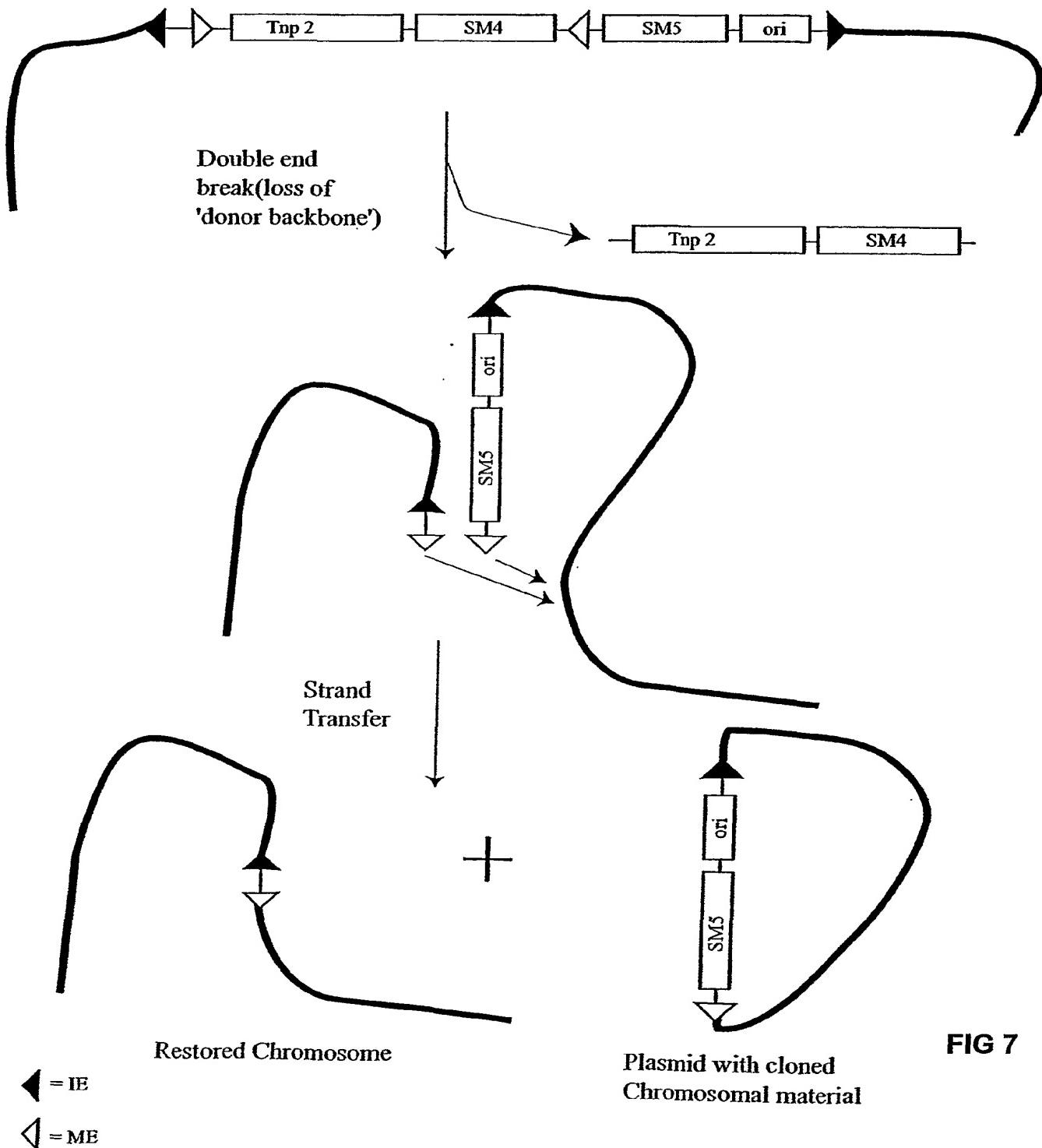
 $SM4^r + SM5^r =$  No transposition $SM4^s + SM5^r =$  Transposition event: Inversion or 'wrong' deletion $SM4^s + SM5^s =$  Transposition event: Deletion

FIG 6

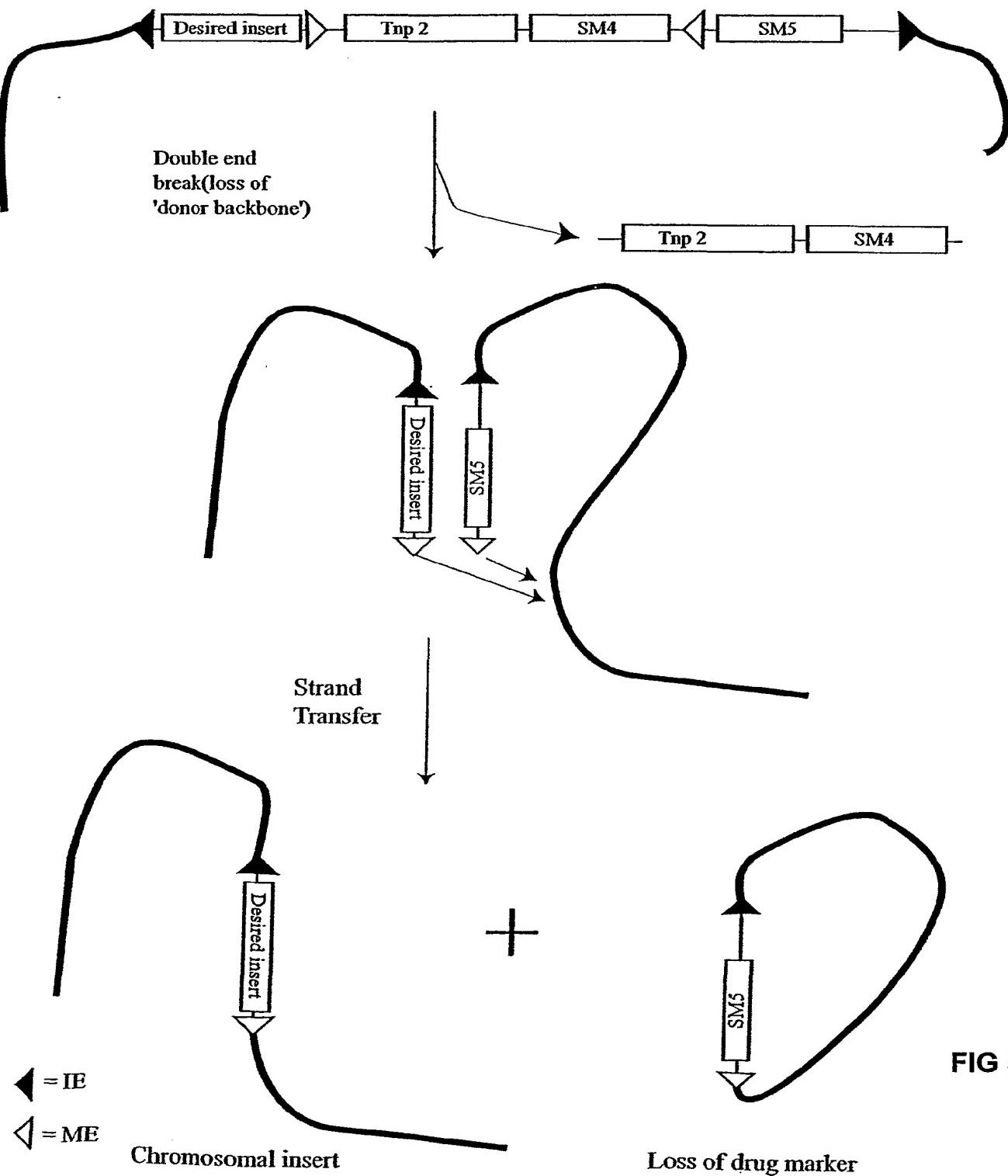
7/10

## Direct cloning of chromosomal DNA.

**FIG 7**

8/10

## Insertion of DNA and removal of drug marker.

**FIG 8**

9/10

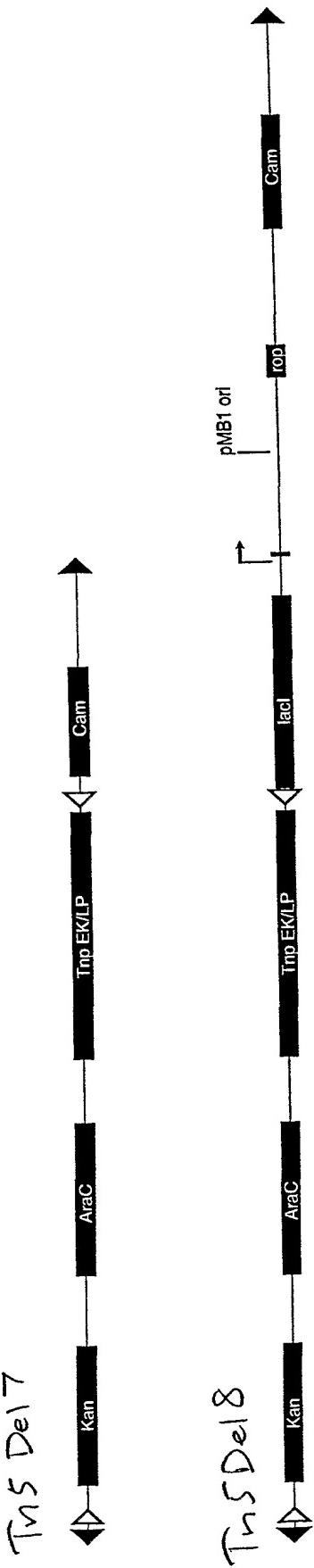
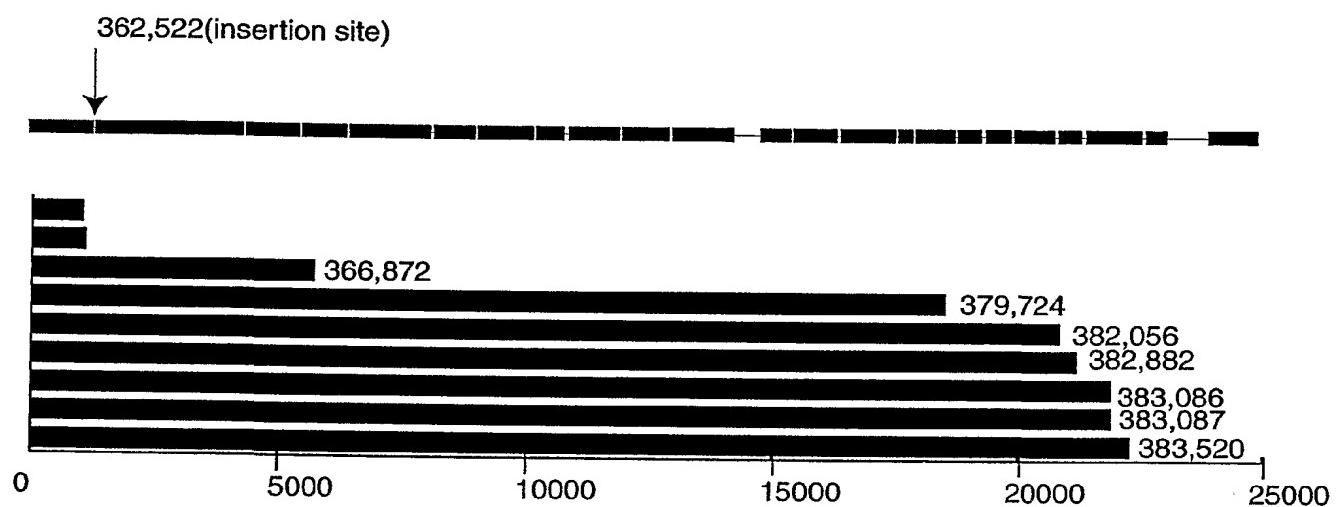


FIG 9

10/10

**FIG 10**

## SEQUENCE LISTING

<110> Goryshin, Igor Y  
Naumann, Todd A  
Reznikoff, William S

<120> DOUBLE TRANSPOSITION METHODS FOR MANIPULATING NUCLEIC ACIDS

<130> 960296.97541

<140>

<141>

<150> 60/251482  
<151> 2000-12-05

<160> 10

<170> PatentIn Ver. 2.1

<210> 1  
<211> 38  
<212> DNA  
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: LINKER A (FULL LENGTH)

<400> 1

ctgtctcttg atcagatcta ctttgtata agagtcag

38

<210> 2  
<211> 32  
<212> DNA  
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: LINKER A (COMPRESSED)

<400> 2

ctgtctcttg atcagatgtg tataaggatc ag

32

<210> 3  
<211> 38  
<212> DNA  
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: LINKER B (FULL LENGTH)

<400> 3

ctgtctcttg atcagatcta gatgtgtata agagacag

38

<210> 4  
<211> 32  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: LINKER B  
(COMPRESSED)

<400> 4  
ctgtctcttg atcagatgtg tataagagac ag 32

<210> 5  
<211> 24  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: PRIMER FWD2

<400> 5  
cagatctcat gcaagcttga gctc 24

<210> 6  
<211> 25  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: PRIMER

<400> 6  
ggtgtcgcttt ctgacaaaact cgggc 25

<210> 7  
<211> 25  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: PRIMER

<400> 7  
acgcgaaaata cgggcagaca tggcc 25

<210> 8  
<211> 3277  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: Pgt4

<400> 8  
gacagctgatc tcttgatcag atctcatgca agcttggctg cagggggggg gggaaaggcca 60

cgttgtgtct caaaaatctct gatgttacat tgccacaagat aaaaatataat catcatgaac 120  
aataaaaactg tctgcttaca taaacagtaa tacaagggggt gttatgagcc atattcaacg 180  
gaaaacgtct tgctcgagggc cgcgattaaa ttccaacatg gatgtgtatt tatatggta 240  
taaatgggt cgcgataatg tcgggcaatc aggtgcgaca atctatcgat tgtatggaa 300  
gcccgatgct ccagagttgt ttctgaaaca tggcaaagggt agcgttgcca atgatgttac 360  
agatgagatg gtcagactaa actggctgac ggaatttatg cctcttccga ccatcaagca 420  
ttttatccgt actccgtatg atgcattgtt actcaccact gcgatccccg gaaaaacagc 480  
attccaggta ttagaagaat atccctgattc aggtgaaaat attgttgatg cgctggca 540  
gttcctgcgc cggttgcatt cgattccgt ttgttaattgt ccttttaaca gcgatcgct 600  
atttcgtctc gctcaggcgc aatcacgaat gaataacggg tttgttgatg cgagtgattt 660  
tgatgacgag cgtaatggct ggcctgtga acaagtctgg aaagaaatgc ataagctttt 720  
gccattctca ccggattcag tcgtcactca tggtgatttc tcacttgata accttatttt 780  
tgacgagggg aaattaatag gttgtattga tggtgacga gtcgaaatcg cagaccgata 840  
ccaggatctt gccatcctat ggaactgcct cggtgagttt ttccttcat tacagaaaacg 900  
gtttttcaa aaatatggta ttgataatcc tgatatgaat aaattgcagt ttcatttgat 960  
gctcgatgag tttttctaat cagaatttgtt taattgggtt taacactggc agagcattac 1020  
gctgacttga cgggacggcg gctttgtga ataaatcgaa cttttgtga gttgaaggat 1080  
cagatcacgc atcttcccga caacgcagac cttccgtgg caaagcaaaa gttcaaaatc 1140  
accaactggt ccacctacaa caaagcttc atcaaccgtt gtccttcac tttctggctg 1200  
gatgatgggg cgattcaggc ctggatgttgc tcagcaacac ctcttcacg aggccagact 1260  
cagcgcggcc ccccccctgc aggtcgactc tagaggatcc cgggttaccg agctcgaatt 1320  
cagatctgt caagagacag ctgtcgacgt caggtggcac ttttggggaa aatgtgcgcg 1380  
gaacccctat ttgttttattt ttcttaataac attcaaataat gtatccgtc atgagacaat 1440  
aaccctgata aatgcttcaa taatattgaa aaaggaagag tatgagttt caacatttcc 1500  
gtgtcgccct tattccctt ttgcggcat ttgccttc tggtttgtt cacccagaaa 1560  
cgctggtaaa agtaaaaatg gctgaagatc agttgggtgc acgagttggg tacatcgaa 1620  
tggatctcaa cagcggtaaatg atccttgaga gtttgcggcc cgaagaacgt ttccaaatga 1680  
tgagcactt taaagttctg ctatgtggcg cgttattatc ccgttattgac gcccggcaag 1740  
agcaactcgg tcgcccgcata cactattctc agaatgactt gtttagtac tcaccagtca 1800  
cagaaaagca tcttacggat ggcatgacag taagagaatt atgcagtgtt gccataacca 1860  
tgagtgataa cactcgccgc aacttacttc tgacaacgtt cggaggaccg aaggagctaa 1920  
ccgctttttt gcacaacatg ggggatcatg taactcgcc ttatcggttgc tgatcggttgg 1980  
tgaatgaagc cataccaaac gacgagcggtt acaccacgtt gcctttagca atggcaacaa 2040  
cggtcgccaa actattaact ggcaactac ttactcttagc ttcccgccaa caattaatag 2100  
actggatgga ggcggataaa gttcaggac cacttctgcg ctggccctt ccggctggct 2160  
ggtttatttc tgataaatct ggagccgggtt agcgtgggtc tcgcgttgc attgcagcac 2220  
tggggccaga tggtaagccc tccctatcg tagttatcta cacgacgggg agtcaggcaa 2280  
ctatggatga acgaaaataga cagatcgctg agataggtgc ctcactgatt aagcatttgtt 2340  
aactgtcaga ccaagtttac tcataatatac tttagattga tttaaaactt catttttaat 2400  
ttaaaaaggat ctatgtgaaatg atcctttttt ataatctcat gacaaaatc ccttaacgtg 2460  
agttttcgat ccactgagcg tcagaccccg tagaaaaagat caaaggatct tcttgagatc 2520  
cttttttctt ggcgttaatc tgctgtttgc aaacaaaaaaa accaccgtt ccagcggtgg 2580  
tttgggttgc ggtcaagag ctaccaactc ttttccgaa ggttaactggc ttcagcagag 2640  
cgccgatacc aaatactgtc cttcttagtgc agccgttagtt aggccaccac ttcaagaact 2700  
ctgttagcacc gcctacatac ctgcgtctgc taatctgtttt accagtggct gctgccagt 2760  
gcccataagt gttcttacc ggggttgact caagacgata gttaccggat aaggcgcagc 2820  
ggtcgccgtt aacgggggggt tcgtgcacac agcccgactt ggagcgaacg acctacaccg 2880  
aactgagata cctacagcggtt gagctatgag aaagcgccac gcttcccgaa gggagaaaagg 2940  
cgccgatacc tccggtaagc ggcagggtcg gaacaggaga gccgcacgagg gagcttccag 3000  
ggggaaaacgc ctggtatctt tatagtcctg tcgggtttcg ccacctctga ttgagcgtc 3060  
gatttttgtt atgctcgtca gggggggcgga gcctatggaa aaacgcccagc aacgcggct 3120  
ttttacgggtt cctggccctt tgctggccctt ttgttcacat gttcttccctt gctgttatccc 3180  
ctgattctgt ggataaccgtt attaccgcct ttgagtgagc tgataaccgtt cgccgcagcc 3240  
gaacgaccga ggcgagcgag tcagtgagcg aggaagc 3277

<210> 9  
<211> 7814  
<212> DNA  
<213> Artificial Sequence

&lt;220&gt;

&lt;223&gt; Description of Artificial Sequence: Pgt7

<400> 9  
gacagctgtc tcttgatca gatctcatgca agcttcagct cactcactca agatgtgtat 60  
aagagacagt cgagatcccc gccacggttt atgagagctt tggtttaggt ggaccagttg 120  
gtgatttga acttttgctt tgccacggaa cggctcgct tgcggaaatg atgcgtgatc 180  
tgatccttca actcagcaaa agttcgattt attcaacaaa gcccgtcc cgtcaagtca 240  
gcgtaatgtc ctgccagtgt tacaaccaat taaccaattc tgattagaaa aactcatcga 300  
gcatcaaataa aactgtcaat ttattccatat caggattata aataccatat ttttggaaaa 360  
gccgtttctg taatgaagga gaaaactcac cgaggcagtt ccataaggatg gcaagatct 420  
ggtatcggtc tgcattccg actcgccaa catcaataca acctattat ttcccctcg 480  
caaaaataaag gttatcaagt gagaatcac catgagtgc gactaatcc ggtgagaatg 540  
gcaaaaatgtt atgcatttttctt ccagcactt gttcaacagg ccagccattt cgctcgat 600  
caaaaatcaact cgcataacc aaaccgttat tcattcgta ttgcgcctga gcgagacgaa 660  
atacgcgatc gctgttaaaa ggacaattac aaacaggaaat cgaatgcac cggcgcagga 720  
acactgcccgg cgcataaca atattttacat ctgaatcagg atatttttctt aataccttgg 780  
atgctgtttt tccggggatc gcagtggta gtaaccatgc atcatcaggatg tacggataa 840  
aatgcttgat ggtcggaaga ggcataaaat ccgtcagcca gtttagtctg accatctcat 900  
ctgttaacatc attggcaacg ctaccttgc catgtttcag aaacaactct ggcgcatecg 960  
gcttccata caatcgatag attgtcgac ctgattgccc gacattatcg cgagccattt 1020  
tataacccata taaatcagca tccatgttgg aatttaatcg cggccctcgag caagacgtt 1080  
cccggttgaat atggctcata acaccccttg tattactgtt tatgtaaagca gacagtttta 1140  
ttgttcatga tgatataattt ttatcttgcgtaa caatgtaaaca tcagagattt tgagacacaa 1200  
cgtggcttc ccccccccccc ctatgcggta taaaataccg cacagatgcg taaggagaaa 1260  
ataccgcata aggacgcgcctt ctgtagcggc gcattaagcg cggcgggtgt ggtggttacg 1320  
cgcagcgtga cgctacactt gccagcgcctt tagcgcgcgc tcctttcgct ttcttccctt 1380  
cctttctcgc cacgttcgcctt atgcataatg tgcctgtcaa atggacgaaag cagggattct 1440  
gcaaacccta tgctactccg tcaagccgtc aattgtctga ttctgttacca attatgacaa 1500  
cttgacggct acatcattca ctttttcttc acaacccggca cggaaactcgc tcgggctggc 1560  
cccggtgcat ttttaataa cccgcgagaa atagagtgtt tcgtcaaaac caacattgcg 1620  
accgacgggtg gcgataggca tccgggttgtt gctcaaaagc agcttcgcctt ggctgataacg 1680  
ttggtcctcg cgccagctta agacgctaattt cccttaactgc tggcgaaaaa gatgtgacag 1740  
acgcgcacggc gacaagcaaa catgctgtgc gacgctggcg atatcaaaat tgctgtctgc 1800  
caggtgatcg ctgatgtact gacaaggctc gcgtacccga ttatccatcg gtggatggag 1860  
cgactcgatc atcgcttcca tgcgcgcgcg taacaattgc tcaaggatgtt ttatcgccag 1920  
cagctccgaa tagcgcctt ccccttgcggc ggcgttaatg atttgcacaa acaggtcgat 1980  
gaaatgcggc tgggtgcgtt catccggcg aaagaacccc gtattggcaa atattgacgg 2040  
ccagttaaagc cattcatgcc agtagggcg cgacgaaag taaacccact ggctgatacca 2100  
ttcgcgagcc tccggatgac gaccgtatgtt atgaatctct cctggcgaaa acagcaaaat 2160  
atcacccggc tggcaaaacaa attctcgatc ctgatttttcc accacccctt gaccgcgaat 2220  
ggtgagattt agaatataac ctttcattcc cagcggtcgg tcgataaaaaa aatcgagata 2280  
accgttggcc tcaatcgccg taaaacccgc caccagatgg gcattaaacg agtatcccg 2340  
cagcaggggc tcatttgcg cttcagccat acttttccata ctccgcctt tcagagaaga 2400  
aaccaattgt ccatattgc tcaagacattt ccgtcactgc gtctttact ggcttcttc 2460  
gctaaccaaa ccgttaaccc cgcttattaa aagcattctg taacaaagcg ggaccaaaagc 2520  
catgacaaaa acgcgtaaacaa aagtgtcta taatcaccgc agaaaagtcc acattgatta 2580  
tttgcacggc gtcacacttt gctatccat agattttttca tccataagat tagcggatcc 2640  
tacctgacgc tttttatcgcc aactctctac tgggttcacca taccgggtttt tttggcttag 2700  
aaataattttt gtttaactttt aagaaggaga tataaccatg ataacttctg ctcttcattcg 2760  
tgcggccgac tggctaaat ctgtgttctc ttccggcgcc ctgggtgatc ctgcgcgtac 2820  
tgcccgctt gtttaacgtcg ccggccaaattt gcaaaaatat tctggtaaat caataaccat 2880  
ctcatcagag ggttagtaaag ccggccagga aggcgttac cgatttatcc gcaatcccaa 2940  
cgtttctgccc gaggcgtca gaaaggctgg ccgcattgcac acagtcaagt tggctcagga 3000  
gtttccgaa ctgctggcca ttgaggacac cacctcttgc agttatcgcc accaggtcgc 3060  
cgaagagctt ggcaagctgg gctctattca ggataaaatcc ccgcggatggt gggttcactc 3120  
cggtcttgcg ctcgaggccca ccacattccg caccgttagga ttactgcattc aggagtggtg 3180  
gatgcgcccgg gatgaccctg ccgatgcggta tggaaaggag agtggcaaat ggctggcagc 3240  
ggccgcaact agccggttac gcatggcagc catgatgac aacgtgattt cggctctgtga 3300

ggcgcgaagcc gatattcatg ctatctgca ggacaaactg ggcataacg agcgcttcgt 3360  
ggtgcgtcc aagcacccac gcaaggacgt agagtctggg ttgtatctgt acgaccatct 3420  
gaagaaccaa cggagttgg gtggatatca gatcagcatt ccgaaaagg gcgtgggga 3480  
taaacgcggt aacgtaaaa atcgaccagc ccgcaaggcg agctttagcc tgcgcagtgg 3540  
gcccgcgtc ctaaaacagg ggaatatcac gctcaacgcg gtgtggccg aggagattaa 3600  
cccgcggaaag ggtgagaccc cggtgaaatg gtgttgcgt accagcgaac cggtcagtc 3660  
gctagccaa gccttgcgog tcatcgacat ttataccat cgctggcga tcgaggagtt 3720  
ccataaggca tgaaaaaccc gaggcaggagc gagaggccaa cgcatggagg agccggataa 3780  
tctggagccg atggtctcgta tcctctcggt tggtcggtc aggctgttac agctcagaga 3840  
aagcttcacg cccgcgaag cactcaggc gcaaggctg ctaaaggaaag cggaacacgt 3900  
agaaagccag tcccgagaaa cggtgctgac cccggatgaa tgtagctac tggcttatct 3960  
ggacaaggaa aaacgcgaagc gcaaaagagaa agcaggtagc ttgcagtggg cttacatggc 4020  
gatacgtaga ctggcggtt ttatggacag caagcgaacc ggaattgcca gctggggcgc 4080  
cctctggtaa ggttgggaag ccctgcaaag taaaactgat ggcttcttg ccgccaagga 4140  
tctgtatggcg caggggatca agatctgatc cgggcttcc ccgtcaagct otaaatcggg 4200  
ggctcgactg tctcttatac acatcttgc tgagtggaaa cctgcattaa tgaatcgggt 4260  
accgagctcg aattacttca ctgacaccct catcagtgcc aacatagtaa gccagtatac 4320  
actccgctag cgctgatgtc cggcggtgct tttgcgtt ccgaccaccc cgtcagtagc 4380  
tgaacaggag ggacagctga tagaaacaga agccactggc gcacacctaaa aacaccatca 4440  
tacactaaat cagtaagtt gcagcatcac ccgacgcact ttgcggcga taaaatacctg 4500  
tgacggaaga tcaacttcgca gaataaataa ttgcgtt ccctgttgcg 4560  
cctggccaa cttttggcga aatagagacg ttttttttgc gttatcgaga ttttcaggag 4620  
accataatga aataagatca ctaccggcgc ctggatatac caccgttgcg atatccaaat 4680  
ctaaggaagc taaaatggag aaaaaaaatca agtgcgttcc tcaatgtacc tataaccaga 4740  
ggcatcgtaa agaacatttt gaggcatttc agaccgtaaa gaaaaataaag cacaagttt 4800  
ccgttcaagct ggatattacg gccttttaa ttttttttgc tcatccggaa ttccgttatgg 4860  
atccggcctt tattcacatt ctggccgc ttttttttgc ttttttttgc ttttttttgc 4920  
caatgaaaga cggtgagctg gtgatatggg atagtgtca ccctgttac accgtttcc 4980  
atgagcaaac tggaaacgttt tcatacgctt ggagtgaata ccacgcgtt ccctggcagt 5040  
ttctacacat atattcgcaa gatgtggcgt gttacgggtga aaacctggcc tattttccct 5100  
aagggtttat tgagaatatg ttttttttgc cagcaatcc ctgggtgagt ttccaccagg 5160  
ttgatttaaa cgtggccaaat atggacaact tcttcgcccc cgttttcacc atggcaaat 5220  
attatacgca aggcgacaag gtgcgtatgc ggagtgaata ccacgcgtt ccctggcagt 5280  
gtgatggctt ccatgtcggc agaatgccta atgaattaca acagtactgc gatgagtggc 5340  
aggcggggc gtaattttt taaggcagtt attgggtgccc taaaacgcct ggtgcacgc 5400  
ctgaataatt gataataaagc ggtatgttgc gtttatgtct atgggtggtt taccgttta 5460  
gtcgggttcag ggcagggtcg taaaatagcc ttctcaaatttgc cctgaggccca gtttgcctc 5520  
ttgactaccg gaagcagttt gaccgtgtcc atgatcattt attctgcctc ccagagcctg 5580  
gctctccccg tggaggtaaat aatttgcgtt atgttaggtt tccacagggt agccagcagc 5640  
ataaaaaacgg ttagcgcttc gtaatacag tgcaggccgc ttgtttccggc gtgggtatgg 5700  
atcctcgcat gcaatccgg aacataatgg ggactgttgg ggcgtccgc caccgttcc acgagttgc 5760  
tggcaggccc cgtggccggg ggactgttgg cggcgtccgc atcgtatcaaa gagacagctg 5820  
tgataaaagaa gacagtcatc agtgcggcga cggaaatttgc atcgtatcaaa gagacagctg 5880  
tcgacgtcg gtggacttt tcggggaaat gtgcggaa ccccttatttgc ttttttttgc 5940  
taaatacatt caaatatgtc tccgtcatg agacaataac cctgataaaat gcttcaataa 6000  
tattaaaaaa ggaagagttt gaggatttgc catttccgtg tcgccttgc tccctttttt 6060  
gcggcatttt gccttcgtt ttttgcgttcc ccagaaacgc tggtaaaatg aaaagatgt 6120  
gaagatcagt tgggtgcacg agtgggttac atcgaactgg atctcaacag cggtaagatc 6180  
cttgagagtt ttcgccccca agaacgtttt ccaatgtatc gcaacttttgc agttctgcata 6240  
tgtggcgcgg tattatcccg tatttgcgttcc gggcaagagc aactcggtcg ccgcatacac 6300  
tattctcaga atgacttgcgt tgactactca ccagtcacag aaaagcatct tacggatggc 6360  
atgacagtaa gagaattatg cagtgtgtcc atcgaactgg atctcaacag cggtaagatc 6420  
ttacttctga caacgtatcg aggaccgaag ccaatgtatc gcaacttttgc agttctgcata 6480  
gatcatgtaa ctgccttgc tcgttggaa cggcgtcatc gggcaagagc aactcggtcg ccgcatacac 6540  
gagcgtgaca ccacgtatgc tgcgttgcgttcc gcaacaacgt tgcgttgcgttcc attaactggc 6600  
gaactactta ctctagcttc ccggcaacaa gggcaacaa cggcgtcatc gggcaagagc tttatagact ggtatggaggc ggataaaagtt 6660  
gcaggaccac ttctgcgtc ggccttcgg gctggctgttcc gcaacaacgt tgcgttgcgttcc attaactggc 6720  
gccgggtgagc gtgggtctcg cggatcatttgc gcaacaacgt tgcgttgcgttcc attaactggc 6780  
cgtatcgtag ttatctacac gacggggagt caggcaacta tgatgtacgc aaatagacag 6840  
atcgctgaga taggtgcctc actgattaag cattggtaac tgcgttgcgttcc gcaacaacgt tgcgttgcgttcc attaactggc 6900

tatatacttt agattgattt aaaacttcat ttttaattta aaaggatcta ggtgaagatc 6960  
 cttttgata atctcatgac caaaaatccct taacgtgagt ttgcgttcca ctgagcgtca 7020  
 gaccggtag aaaagatcaa aggatcttot tgagatcctt ttttctgcg cgtaatctgc 7080  
 tgctgc当地 caaaaaacc accgctacca gcgggtggttt gttgccgaa tcaagagcta 7140  
 ccaactctt ttccgaaggta aactggcttc agcagagcgc agataccaa tactgtcctt 7200  
 ctatgttagc cgtagttagg ccaccacttc aagaactctg tagcaccgccc tacataccct 7260  
 gctctgctaa tcctgttacc agtggctgct gcagtgccgataa gttcgatg tcttaccggg 7320  
 ttggactcaa gacgatagtt accggataag ggcgcagcggt cgggctgaac ggggggttcg 7380  
 tgcacacagc ccagcttggaa gcgaacgacc tacaccgaa tgagataacct acagcgtgag 7440  
 ctatgagaaa gcgccacgct tcccgaaggg agaaaggcgg acaggtatcc ggtaaagcggc 7500  
 agggtcggaa caggagagcgc cacgaggag cttccagggg gaaacgcctg gtatctttat 7560  
 agtcctgtcg ggttcgcca cctctgactt gagcgtcgat tttgtgtatg ctcgtcaggg 7620  
 gggcggagcc tatggaaaaaa cgccagcaac gcggcccttt tacggttcct ggcctttgc 7680  
 tggccttttgc ctcacatgtt cttccgtcg ttatcccctg attctgtgga taaccgtatt 7740  
 accgcctttg agtgagctga taccgctcgcc cgagccgaa cgaccgagcgc cagcgagtca 7800  
 gtgagcggagg aagc 7814

&lt;210&gt; 10

&lt;211&gt; 9265

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Description of Artificial Sequence: Pgt8

&lt;400&gt; 10

ctaagaaacc attattatca tgacattaac ctataaaaat agggttatca cgaggccctt 60  
 tcgtcttcaa gggtcgtca ctgcccgtt tccagtcggg aaacctgtcg tgccagctgc 120  
 attaatgaat cggccaacgc gcggggagag ggggtttgcg tattggcgc cagggtgttt 180  
 tttctttca ccagttagac gggcaacagc tgattgcctt tcaccgcctg gcccctgagag 240  
 agttgcagca agcggtccac gctggtttgc cccagcaggc gaaaatccctg tttgtatgt 300  
 gttgacgccc ggtatataaca tgagctgtct tccgtatcgt cgtatcccac taccgagata 360  
 tccgcaccaa cgcgcagccc ggactcggtt atggcgcgcg ttgcgcggcag cccatctga 420  
 tcgttggcaa ccagcatcgc agtggaaacgc atgcccctcat tcagcatttgc catgtttgt 480  
 taaaaaccgg acatggact ccagtgcct tcccggtccg ctatcggtcg aatttgatttgc 540  
 cgagttagat atttatgcca gccagccaga cgcagacgc cggagacaga acttaatggg 600  
 cccgctaaca ggcgcatttgc ctggtagcccaatgcgacca gatgtccac gcccagtcgc 660  
 gtaccgtctt catggagaa aataatactg ttgtatgggtg tctggtcaga gacatcaaga 720  
 aataacgccc gaacattagt gcaggcagct tccacagcaa tggcatccctg gtcatccagc 780  
 ggatagttaa tgatcagccc actgacgcgt tgccgcgagaa gattgtgcac cggcgctta 840  
 caggcttcga cgcgcgttcg ttctaccatc gacaccacca cgctggcacc cagttgatcg 900  
 ggcgcgagatt taatcgccgc gacaatttgc gacggcgcgt gcagggccag actggaggtg 960  
 gcaacgc当地 tcagcaacga ctgtttgccc gccagttgtt gtgcacgcg gttggaaatg 1020  
 taattcagct cccgcattcgc cgcttccact ttttccgcg ttttcgcaga aacgtggctg 1080  
 gcctggtca ccacgcgggaa aacggctcga taagagacac cggcataactc tgacatcg 1140  
 tataacgtta ctgtttcac attcaccacc ctgaaatttgc tctctccgg ggcgtatcat 1200  
 gcccattaccgc gaaagggtttt gcaccattcg atgggttcgg cagcgttggg tcctggccac 1260  
 ggggtgc当地 gatgtgtcgc ctgtcggtga ggacccggct agagtcgca cgc当地 1320  
 gtgagtttagc tcactcatta ggcacccagg gctttacact ttatgtcttcc ggctcgatcg 1380  
 ttgtgtgaa ttgtgagccg ataacaattt cacacaggaa acaggggtac cctgcttgc 1440  
 aacaaaaaaaaa ccaccgtac cagcgggtgtt ttgtttgccc gatcaagagc taccactt 1500  
 ttttccgaag gtaactggct tcagcagacgc gcaagatcca aataactgtcc ttcttagt 1560  
 gcccgtatc ggccaccact tcaagaactc tgcacccggc cctacatacc tcgctctgct 1620  
 aatcctgtta ccagtggctg ctgcccgtgg cgataagtgc tgcgttaccg ggttggactc 1680  
 aagacgatag ttaccggata aggccgacgc gtcgggctga acgggggggtt cgtgcacaca 1740  
 gcccagctt gaggcaacga cctacaccga actgagatac ctacacgcgtg agctatgaga 1800  
 aagcgc当地 cttccggaa gggaaaaggc ggagaaaggc ggacaggat cccgtaaagcgc gcaagggtcgg 1860  
 aacaggagag cgcacgagg agcttccagg gggaaaacgcc tggtatctt atagtcctgt 1920  
 cgggttccgc caccctctgac ttgagcgtcg atttttgta tgctcgtag gggggcggag 1980

cctatggaaa aacgccagca acgcggcctt tttacggttc ctggcctttt gctggccttt 2040  
tgctcacatg ttcttcctg cgttatcccc tgattctgtg gataaccgta ttaccgcctt 2100  
tgagttagct gataccgctc gccgcagccg aacgaccgag cgcagcgagt cagtgagcga 2160  
ggaagcggaa gagcgcctga tgcggattt ttccttacg catctgtgcg gtatttcaca 2220  
ccgcataatgg tgcactctca gtacaatctg ctctgatgcc gcatagttaa gccagtatac 2280  
actccgctat cgctacgtga ctgggtcatg gctgcgc(ccc gacacccgcc aacacccgct 2340  
gacgcgcctt gacgggcttg tctgctccc gcacatccgctt acagacaaggc tgtgaccgtc 2400  
tccgggagct gcatgtgtca gaggtttca ccgtcatcac cgaaacgcgc gaggcagctg 2460  
cggtaaagct catcagcgtg gtcgtgaagc gattcacaga tgtctgcctg ttcatccgcg 2520  
tccagctcgt tgagtttctc cagaagcggt aatgtctggc ttctgataaa gcgggcccattg 2580  
ttaaggcggg tttttcctg tttggtcaact gatgcctccg tgtaaaggggg aattctgttc 2640  
atgggggtaa tgataccgt gaaacgagag aggatgctca cgataacgggt tactgatgat 2700  
gaacatgccc ggtaacttggg acgttgtgag ggtaaacaac tggcggatg gatgcggcgg 2760  
gaccagagaa aaatcactca gggtaatgc cagcgcctcg ttaatacaga tgttaggtgtt 2820  
ccacagggtt gccagcagca tcctgcgtt cagatccggg acataatggt gcagggcgct 2880  
gacttcgcg tttccagact ttacgaaaca cgaaacccga agaccattca tgggtttgtc 2940  
caggtcgcag acgttttgcg gcagcagtgc cttcacgttc gtcgcgtat cggtgattca 3000  
ttctgctaacc cagtaaggca accccgcctt cctagatccct tttagcttta tgcttgtaaa 3060  
ccgttttgcg aaaaaatttt taaaataaaa aaggggaccc ttagggtccc caattaatta 3120  
gtaatataat ctattaaagg tcaattaaaaa ggtcatccac cggtatccgag ctcgaattgt 3180  
aagaggttcc aactttcacc ataatgaaaat aagatcacta ccgggcgtat ttttgagtt 3240  
atcgagattt tcaggagctt aggaagctaa aatggagaaa aaaatcactg gatataccac 3300  
cggtgatata tcccaatggc atcgtaaaga acattttgcg gcatttcgtt cagttgtca 3360  
atgtacctat aaccagaccc ttcagctggg tattacggcc tttttaaaaga ccgtaaagaa 3420  
aaataaagcac aagttttatc cggccttat tcacattctt gcccgcctga tgaatgctca 3480  
tccggaaattt cgtatggcaa tgaagacgg tgagctgggt atatggata gtgttcaccc 3540  
ttgttacacc gttttccatg agcaaactga aacgttttca tcgctctggg gtgaataacca 3600  
cgacgatttcc cggcagtttcc tacacatata ttcgcaagat gtggcgtt acggtgaaaa 3660  
cctggcctat ttcctttaaag ggtttattga gaatatgtt ttctgtctcag ccaatccctg 3720  
ggtgagtttcc accagttttt atttaaacgt gccaatatg gacaacttct tcgccccctg 3780  
tttcaccatg ggcaaatatc atacgcaagg cgacaagggtt ctgatgccgc tggcgattca 3840  
ggttcatcat gccgtctgtg atggcttcca tgcggcaga atgctaattt aattacaaca 3900  
gtactgcgtt gagtggcagg gcggggcgta attttttaaa ggcagttatt ggtgcctta 3960  
aacgcctgtt gctacgcctg aataattgtt aataagcgga tgaatggcag aatttcgaaa 4020  
gcaaatttca cccggctgtc gtttcaggcc agggtcgtt aatagccgt tatgtctatt 4080  
gctggtttac cgggttattt actaccggaa gcagtgtgac cgtgtcttc tcaaatgcct 4140  
gaggccagg tgcgtcaggct ctccccgtt agtaataat tgacgatatg atcattttatt 4200  
ctgcctccca gagcctgata aaaacgttta ggccttcgtt aatacagatg taggtgttcc 4260  
acagggttagc cagcagcattt ctgcgtatgc gatccggaaac ataatggtgc agggcgctt 4320  
tttcggcgtt ggtatgggtt caggccccgtt gggccggggg ctgttggcgc ctgcccggac 4380  
ctgtcctacg agttgcattt taaagaagac agtcataatgtt gcccgcacga aattcagatc 4440  
tgatcaagag acagctgtcg acgtcagggtt gcaacttttcg gggaaatgtt cgcggaaaccc 4500  
ctatttgcattt attttctaa atacatttca atatgtatcc gctcatgaga caataaccct 4560  
gataaaatgtt tcaataatat taaaaaaaggaa agagtatgag tattcaacat ttccgtgtcg 4620  
cccttatttcc cttttttcg gcatttttgc ttcctgtttt tgctcacccca gaaacgcctgg 4680  
tgaaaatggaa agatgtctttt gatcagggtt gtcacacgg gagcttccag ggggaaacgc 4740  
ctggtatctt tatagtctgtc tcgggtttcg ccacctctga cttgagcgtc gattttgtt 4800  
atgcgtgtca gggggggcgaa gcctatggaa aaacgccttgc aacgcggccctt ttttacgggtt 4860  
cctggcctt tgcgtggcctt ttgctcacat gttctttctt gctttatccc ctgattctgt 4920  
ggataaccgtt attaccgcctt ttgagtgagc tgataccgtt cggccgcagcc gaaacgaccga 4980  
gcgcagcgag tcagtggatc aggaagcgac agctgtcttct tgcgtatgc tcatgcac 5040  
ttcagcttac tcactcaaga tgcgtataag agacagtgcg gatccccggcc acgggttgatg 5100  
agagctttgt tgcgtggatc ccagttgggtt attttgcact tttgcttgc cacggaaacgg 5160  
tctgcgttgc cgggaagatgc cgtgtatgc tccttcaact cagcaaaagt tcgattttatt 5220  
caacaaagcc gccgtcccgat caagtccgcg taatgtctgtt ccagtgttac aaccaattaa 5280  
ccaatttctga ttagaaaaac tcatcgagca tcaaaatggaa ctgcaattta ttcatatcag 5340  
gattatcaat accatatttt taaaaaaaggcc gtttctgttt tgaaggagaa aactcaccga 5400  
ggcagttcca taggtggca agatccgtt atcggctctgc gattccgact cgtccaacat 5460  
caatacaacc tattaatttcc cctcgtaaa aaataagggtt atcaagtgggaaatcaccat 5520  
gagtgacgac tgaatccggat gagaatggca aaagtttgcatttcc cagacttgc 5580

caacaggcca gccattacgc tcgtcatcaa aatcactcgc atcaacaaa ccgttattca 5640  
ttcgtgattg cgcttgagcg agacaaata cgcgatcgct gttaaaagga caattacaaa 5700  
caggaatcga atgcaaccgg cgccagaaca ctgccagcg atcaacaata tttcacctg 5760  
aatcaggata ttcttctaatt acctggaatg ctgttttcc gggatcgca gtggtagta 5820  
accatgcattc atcaggagta cggataaaat gcttgatggt cggaagaggc ataaattccg 5880  
tcagccagtt tagtctgacc atctcatctg taacatcatt ggcaacgcta ccttgcct 5940  
gtttcagaaa caactctggc gcacatggct tcccatacaa tcgatagatt gtgcacactg 6000  
attgcccac attatcgca gcccatttat acccatataa atcagcatcc atgttggaa 6060  
ttaatcgccg cctcgagcaa gacgttccc gtgaaatatg gctcataaca ccccttgtat 6120  
tactgtttat gtaagcagac agtttattt ttcatgatga tatattttt tcttgcgaa 6180  
tgtaacatca gagatttga gacacaacgt ggotttcccc ccccccccta tgccgtgtga 6240  
aataccgcac agatcgtaa ggagaaaata cgcacatcagg acgcgcctg tagcggcgca 6300  
ttaagcggc cgggtgtggt gttacgcgc acgtgacgc tacacttgcc agcgccttag 6360  
cgcccgctcc ttgcgtttc ttcccttc ttctcgccac gttcgccatg cataatgtgc 6420  
ctgtcaaatg gacaaagcag ggattctgca aaccctatgc tactccgtca agccgtcaat 6480  
tgtctgattt gttaccaatt atgacaactt gacggctaca tcattcaattt tttcttcaca 6540  
accggcacgg aactcgctcg ggctggcccc ggtgcatttt ttaaataaccc gcgagaataa 6600  
gagttgatcg tcaaaaaccaa cattgcgacc gacgggtggcg ataggcatcc ggggtgggtct 6660  
caaaaggcgc ttccctggc tgatacgtt gtcctcgcc cagcttaaga cgctaattccc 6720  
taactgtgg cggaaaagat gtgacagacg cgacggcgac aagcaaacat gctgtcgac 6780  
gctggcgat tcaaaaattgc tgtctgccc gtgatcgctg atgtactgac aagcctcg 6840  
tacccgatta tccatcggtg gatggacgc ctcgttaatc gttccatgc gccgcagtaa 6900  
caattgtca agcagattt tgccagcag ctcgcataatg cgcccttcccc cttgcccggc 6960  
gttaatgatt tgccaaaca ggtcgctgaa atgcggctgg tgcgcttcat ccggcgaaaa 7020  
gaaccccgta ttgcaaaata ttgacggcca gttaaagccat tcatgcccagt aggccgcgg 7080  
acgaaaagtaa acccaactggt gataccattt cgcggctcc ggatgacgc cgtagtatg 7140  
aatctctcct ggccggaaaca gcaaaatatac acccggtcg gaaacaaattt ctcgtccctg 7200  
attttcacc accccctgac cgcgaatggt gagattgaga atataaccc ttattccctc 7260  
cggtcggtcg ataaaaaaaat cgagataacc gttggcctca atcggcgat taccggccac 7320  
cagatggcata ttaaacgagt atcccgccag caggggatca ttttgcgctt cagccataact 7380  
tttcataactc ccccaattca gagaagaaac caattgtcca tattgcatca gacattggcg 7440  
tcactgcgtc ttttactggc tcttctcgct aaccaaaccg gtaaccccgcc ttattaaag 7500  
cattctgtaa caaagcggga ccaaagccat gacaaaaacg cgtaacaaaaa gtgtctataa 7560  
tcacggcaga aaagtccaca ttgattattt gcacggcgatc acactttgct atgcctatgc 7620  
attttatcc ataagattag cggatctac ctgacgctt ttatcgcaac tctctactgt 7680  
ttctccatac ccgtttttt gggctagaaa taattttgtt taactttaag aaggagatata 7740  
aaccatgata acttctgctc ttcatcgatc ggccgactgg gctaaatctg tttcttc 7800  
ggcggcgctg ggtgatcctc gccgtactgc cccgttgggt aacgtcgccg cccaaattggc 7860  
aaaatattct ggttaaatcaa taaccatctc atcagagggt agtaaagccg cccaggaagg 7920  
cgcttaccga ttatccgca atcccaacgt ttctgccag gcgatcagaa aggctggcg 7980  
catgcaaaaca gtcaagttgg ctcaggatgt tcccgaactg ctggccattt aggacacac 8040  
ctctttgagt tatcgccacc aggtcgccga agagcttggc aagctggctt ctattcagga 8100  
taaatccgc ggatgggtggg ttcactccgt tctcttgctc gaggccacca cattccgcac 8160  
cgttaggat tgcgtacagg agtgggtggat gcccggat gaccctgcg atgcggatga 8220  
aaaggagat ggcaaatggc tggcagccgc cgcaactagc cggttacgca tgggcagcat 8280  
gatgagcaac gtgattgcgg tctgtgaccg cgaagccgat attcatgctt atctgcagga 8340  
caaactgcgc cataacgagc gttcgtggcgt ggcgttccaaag caccacgcg aggacgtaga 8400  
gtctgggtt gatctgtacg accatctgaa gaaccaaccg gagttgggtg gctatcagat 8460  
cagcattccg caaaaggccg tggtgataa acgcggtaaa cgtaaaaaatc gaccagcccg 8520  
caaggcgagc ttgagcctgc gcagtggcg catcacgcta aaacagggga atatcacgct 8580  
caacgcgtg ctggccgagg agattaaccc gcccagggt gagacccgt tgaaatggtt 8640  
gttgctgacc agcgaaccgg tcgagtcgt agcccaagcc ttgcgtgtca tcgacattta 8700  
tacccatcgc tggcgatcg aggagtccaa taaggcatgg aaaaccggag caggagccga 8760  
gaggcaacgc atggaggagc cgatataatct ggagccgat gtctcgatcc tctcggtt 8820  
tgcggtcagg ctgttacago tcagagaaag cttcacgccc cgcacgcac tcagggcgca 8880  
aggcgtgcta aagaagccg aacacgtaga aagccagtcc gcagaaacgg tgctgacccc 8940  
ggatgaatgt cagctactgg gctatctgaa caaggggaaa cgcaacgcgca aagagaaagc 9000  
agtagctt cagtgccgtt acatggcgat agcttagactg ggcggtttta tggacagacaa 9060  
gcgaaccggaa attgcccagct gggccgcct ctgttaaggt tgggaagccc tgcaaagtaa 9120  
actggatggc tttcttgccg ccaaggatct gatggcgac gggatcaaga tctgatccgg 9180

gctttccccg tcaagctcta aatcggggc tcgactgtct cttatacaca tcttgagtga 9240  
gtgagaacct gcattaatga atcgg 9265

QBMAD\319926.1